

Chapter 1

Introduction

The majority of the stars form in clusters. Carpenter (2000) reports that 50 – 70% of young ($\leq 10Myr$), and 25 – 70% of the relatively old ($\sim 100Myr$), stellar populations were formed in clusters. Porras et al. (2003) and Lada & Lada (2003) find that 80% and 90%, respectively, of the stars form in clusters with more than 100 members. Furthermore, as indicated by Porras et al. (2003), these large clusters represent 22% of the star forming regions. With only 8% of the stars forming in regions with five to 30 members. However, from these large (≥ 100 members) clusters, only less than 7% will survive as bounded clusters when reaching an age of $120Myr$ (Lada & Lada 2003). The remaining of the star forming regions will become unbounded and their stars will freely populate the galaxy. Thus, to understand the general rules that govern how the majority of stars form, as well as the properties of the stars that populate our galaxy, it is crucial to fully decode the formation and evolution of stellar clusters.

In the first decade of this century, numerical simulations of star forming regions have proved to be of paramount importance in the decoding the very early stages of the star formation process (????). For example, Bate et al. (2003) using smooth particle hydrodynamics simulations were able to reproduce the collapse and fragmentation of a large turbulent molecular cloud ($50M_\odot$ within $0.375pc$ radius), which formed about 50 stars.

Spectacular advances have been made in recent years in the field of stellar cluster formation. New statistical theories of star and planet formation have replaced the standard Shu (1987) model of isolated star formation and offer a much more dynamic view of prompt multiple fragmentation driven by MHD supersonic turbulence in molecular clouds (see Fig. 1, and e.g. Bate et al. 2003; Li et al. 2004; Jappsen et al. 2005). Smooth particle hydrodynamics (SPH) calculations are now able to simulate the collapse and fragmentation of large turbulent molecular clouds forming several hundreds of stars (e.g. Bate 2009). However, the outcome of these simulations depend crucially on the physical parameters of the parent molecular cloud (density, turbulence), and if radiative feedback (Bate 2009) and magnetic field (Hennebelle & Teyssier 2008) are taken into account. In particular,

important open issues remain:

Description of:

- Motivation to investigate nearby young open clusters.
- The importance of the IMF
- Current state of knowledge in dynamical simulations.
- The problematic of constraints in dynamical theories.

It must be clear which is the objective

Description of the Nearby open clusters and their properties.

1.1 The DANCe project

- List of open clusters in the DANce project.

- the importance of the pleiades, why we restrict to it.

It must be clear what are the limitations, the boundaries in which the objective will be searched

Description of the current methodologies used to address the question mentioned previously.

- The works of Sarro, Krone-Martins, Malo, Gagne etc. LAcweing

-The advantages and caveats of the previous methodologies.

-It must be clear the necessity of a new perspective

The proposal we made. The use of Bayesian Hierarchical Models. Benefits and issues of BHM.

Description of the advantages of BHM.

-*i* It must be clear that BHM are the best choice.

Description of the practical issues needed to be solved in order to use BHM.

MCMC techinques and PSO.

-*i* It must be clear that MCMC methods are the best option.

Brief descriptions of our results and how they impact our current knowledge.

-*i* It must be clear that we attained the objective: The pleiades velocity, spatial and mass distributions.

Chapter 2

The Pleiades as a benchmark

2.1 Generalities

The ancient greeks named Pleiades to a crowded group of nine stars which they believe share a common origin. These stars were the seven sisters, that together with their parents the titan Atlas and the nymph Pleione, were put in the sky by the god Zeus.

Today, we call the Pleiades cluster not just to the nine stars that made up the original Pleione family, but to a much larger group, which according to [Bouy et al. \(2015\)](#) adds up to ~ 2100 members. This cluster is fairly close to the sun, $\sim 136\text{ pc}$ (according to [Galli et al. 2017](#)), and is also young in galactic scales, with only $\sim 125\text{ Myr}$ ([Stauffer et al. 1998](#)). Since it is located in the solar neighbourhood it has a distinctive angular velocity in the plane of the sky: about -16 mas/yr in right ascension and 20 mas/yr in declination. Also, it has expelled most of its cocoon gas, which gives it an almost null extinction of $A_v = 0.12$ ([Guthrie 1987](#)). These properties make the Pleiades one of the most studied cluster in the history of astronomy¹. Thus making it also the perfect test case of the methodology developed in this work.

As stated in the previous Chapter, the objective of the present study is to obtain the statistical distributions of the distance, position, velocity, luminosity and mass of the Pleiades cluster. Thus, in the following sections I will describe the current knowledge of the Pleiades, concerning these astrophysical quantities.

2.2 The distance to the Pleiades

2.2.1 Measuring distances

In astronomy, measuring distances is a complicated task. Techniques vary according to the distance scale that they aim to measure. The distance ladder is constructed from smaller to larger distances. The first step in this ladder is the distance to the sun. After

¹Probably just after Orion.

that, the distance to the planets and then to the stars. This work deal only with nearby clusters, thus I only focus on measuring distances to these objects.

The most direct way to measure distance to nearby stars is by means of the trigonometric parallax. This is the relative angular displacement, with respect to the far distant stars, that an object suffers in the course of a year. This relative displacement is time dependent and results from the movement of the earth (thus the observer) on its orbit around the sun. The relative displacement is roughly maximal when measurements are taken at opposed points in the earth's orbit, when they are separated by six months. This maximal displacement is called the parallax of the object and is measured in seconds of arc. If the parallax were measured with infinite precision then the distance to the object will be obtained by simply inverting its parallax. By doing so, the distance will be measured in parsecs. This unit gets its name from parallax-second. Thus an object located at a distance of one parsec from the sun shows a parallax of one arc second. The further the object is, the smaller the parallax gets.

However, as any measurement, parallaxes had uncertainties. These uncertainties usually represent, or are a proxy for, the width of the distribution. Thus, nothing prohibits that a certain combination of distance and uncertainty results in a measured parallax with a negative value. The parallax has a non limited continuous distribution.

When transforming parallaxes into distances we may be tempted to take an statistics of the distribution, the mean for example, and just invert it to obtain the distance. This only holds if the statistic correspond to the true value (i.e. the statistic is unbiased). The true value is that in which the uncertainties are negligible. However, because measurements always have uncertainties and almost always they are not negligible, the inversion of the parallax renders an almost unbiased estimate of the distance only for small values of the relative uncertainty ([Lutz & Kelker 1973](#), mention that a reasonable value is below 0.15-0.20). The shape of the parallax distribution (the second and higher order moments) plays an important roll. If we are interested in the distance and we only have the parallax distribution, this distribution must be transformed into that of distances. However, this transformation is not a simple inversion. Several authors have proposed different approaches to the problem of distance determination using parallaxes, see for example [Lutz & Kelker \(1973\)](#); [Bailer-Jones \(2015\)](#); [Astraatmadja & Bailer-Jones \(2016a,b\)](#). The proper way, as [Bailer-Jones \(2015\)](#) points out, consists of inferring the true distances given the observed parallaxes. For that, a prior on the distance must be established. The authors mentioned before describe three different kinds of priors and the methodology needed to infer the true distances. However, going into deeper detail is beyond the scope of this work.

Now, I focus on the particular case of the distance to the Pleiades. The first parallax measurement of the Pleiades was done by [van Leeuwen \(1999\)](#) using the *Hipparcos* data. Later, the same author ([van Leeuwen 2009](#)) refined his sample and obtained a value of

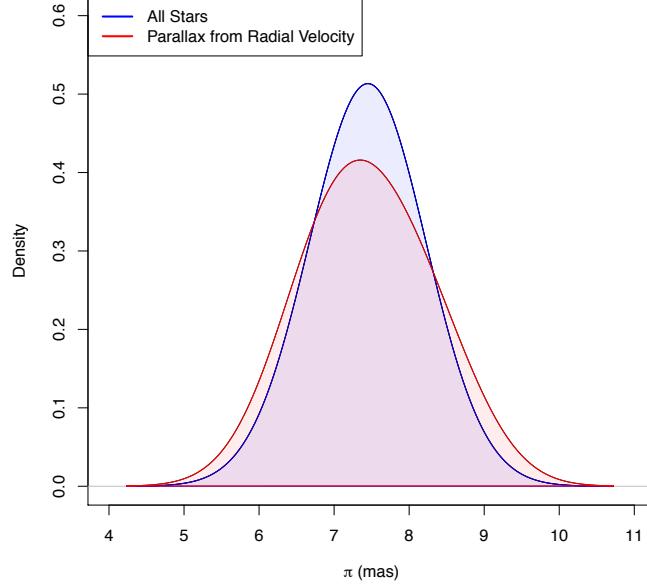
$120 \pm 1.9\text{ pc}$. However, [Gatewood et al. \(2000\)](#) and [Soderblom et al. \(2005\)](#) using also the parallaxes of smaller samples of stars (seven and three, respectively), measured values of $130.9 \pm 7.4\text{ pc}$ and $134.6 \pm 3.1\text{ pc}$, respectively. Finally, [Melis et al. \(2014\)](#) measured $136.2 \pm 1.2\text{ pc}$ using parallaxes of three stars. There is a clear controversy between *Hipparcos* data and that of the rest of the parallax measurements. The current *Tycho-Gaia* data release, TGAS, shows that the Pleiades parallax is $\pi = 7.48 \pm 0.03\text{ mas}$ ([Gaia Collaboration et al. 2017](#)). This seems to indicate that the *Hipparcos* parallaxes were somehow biased. However, this controversy will remain until the totally independent *Gaia* DR2 will be available.

Until this controversy have been solved, I have decided to choose the distance found by our research group, $134.4^{+2.9}_{-2.8}\text{ pc}$ ($7.44 \pm 0.08\text{ mas}$) ([Galli et al. 2017](#)), which is in good agreement with the one found by TGAS. We found this distance using the kinematic parallaxes delivered by the moving cluster technique. This essentially exploits the fact that since clusters are bound, their members show a clear kinematic footprint: they seem to converge to a point in the sky ([Blaauw 1964](#)). Using this point and the velocity of the members (proper motion and radial velocities) it is possible to derive individual parallaxes. Furthermore, these individual parallaxes show a distribution which results from the dispersion of the cluster members along the line of sight. Figures [2.1a](#) and [2.1b](#) show the distribution of parallaxes for the Pleiades candidate members according to [Galli et al. \(2017\)](#) and [Gaia Collaboration et al. \(2017\)](#). As can be seen from these Figures, the results of both works agree on the mean of the parallax distribution. However, they recover different variances. This difference results from the discrepancy in the number of objects (1210 vs. 152) and in the selection function of the two surveys. TGAS sample is limited to the bright objects ($V \sim 11.5$), whereas the Pleiades DANCE DR2 includes the faint end of the distribution ($i \sim 25$).

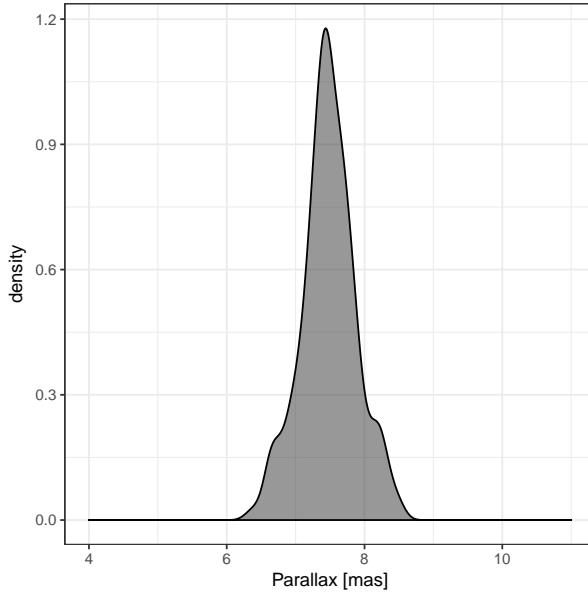
Nevertheless, the parallax distribution is only the depth component of the space distribution of the cluster, the other two components are given by the spatial distribution.

2.3 Spatial Distribution

The spatial distribution is the two dimensional projection in the plane of the sky, the one perpendicular to the line of sight, of the of the cluster three dimensional space distribution. In general, individual object positions in the sphere of the sky, which are commonly known as the coordinates right ascension (R.A.) and declination (Dec.), are fare more easily measured than individual parallaxes. For this reason, just a small fraction of objects has parallaxes. For example, in the case of the Pleiades, only ~ 70 members out of ~ 2100 have *Hipparcos*'s parallaxes. This figure comes after cross matching the candidate members of [Bouy et al. \(2015\)](#) with the *Hipparcos* catalogue ([Perryman et al. 1997](#)). As we saw in the previous section, this figure roughly doubled with the new *Gaia*



(a) Parallaxes according to [Galli et al. \(2017\)](#). The red line shows all their candidate members (1210) while the blue one only those with known radial velocity (64). Reproduced from Figure 13 of [Galli et al. \(2017\)](#), "A revised moving cluster distance to the Pleiades open cluster", A&A, 598.



(b) Parallaxes according to [Gaia Collaboration et al. \(2017\)](#). Only their 152 candidate members.

Figure 2.1: Distribution of parallaxes for the Pleiades members.

DR1 ([Gaia Collaboration et al. 2016](#)). In addition to the scarcity of parallaxes, which is expected to be solved with future *Gaia* data releases, the relative uncertainties in R.A. and Dec. coordinates are far better (10^{-5}) than those of individual parallaxes (10^{-1}).

Due to the previous considerations, the space distribution of the Pleiades has been studied mainly through its spatial distribution. This last has been the subject of several studies. One of the earliest results of the Pleiades spatial distribution was done by [Limber \(1962\)](#). He used a mixture of four indices polytropic distribution, as was described in his earlier [Limber \(1961\)](#) work, to fit the spatial distribution of the 246 candidate members of

Trumpler (1921). These candidates were contained in a 3° radius around *Alcyone* (one of the central most massive stars of the Pleiades cluster).

Later, Pinfield et al. (1998) fitted King profiles (King 1962) of different masses ($5.2, 1.65, 0.83, 0.3 M_\odot$) to candidate members from the literature which covered a 3° radius area. They estimated a tidal radius of 13.1 pc , in which 1194 candidate members were contained. The total mass of these members amounted to $735 M_\odot$. They estimated a mean mass of $0.616 M_\odot$.

On the same year Raboud & Mermilliod (1998) also fitted a King's profile (King 1962) to a list of 270 candidate members contained within a 5° radius area. They found a core radius of 1.5 pc and a tidal radius of 17.5 pc . Using different approaches, they derived a total mass within the range of $500 - 8000 M_\odot$. They also measured an ellipticity of $\epsilon = 0.17$, however they do not make any explicit mention on the position angle of the axis of the ellipse.

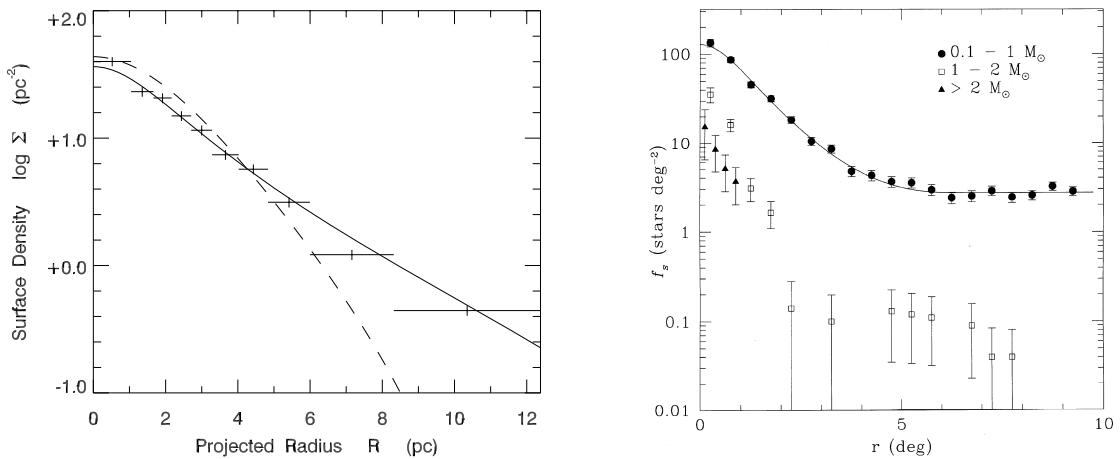
Later, Adams et al. (2001) also fitted a King's profile to objects with membership probabilities $p > 0.3$ within a radius of 10° . They found a core radius of $2.35 - 3.0\text{ pc}$ and a tidal radius of $13.6 - 16\text{ pc}$. They estimate a total mass of $\sim 800 M_\odot$. Their measured ellipticities were in the range $0.1 - 0.35$.

Converse & Stahler (2008) used a sample of 1245 from Stauffer et al. (2007) to fit a King's profile. They obtained a tidal radius of 18 pc and a core radius of 1.3 pc . Later, Converse & Stahler (2010) refined their study and obtained a core radius of $2.0 \pm 0.1\text{ pc}$, a tidal radius of $19.5 \pm 1.0\text{ pc}$ and a total mass of $870 \pm 35 M_\odot$. In Fig. 2.2a, I reproduce the surface density fit obtained by these authors.

The previous summary of results shows at least two interesting points. In first place, King's profile (King 1962) has been the preferred choice for the Pleiades cluster, although it was created to fit the spatial distribution of globular clusters. Since globular clusters are far away than open clusters and in a low density environment, usually the end of their spatial distribution is well within the survey area. The second point concern the increasing trend of the tidal radius with the size of the survey and the publication date. As the surveys increase in area the derived tidal radii increase as well. The exception is the work of Adams et al. (2001). Since they used low membership probability (≥ 0.3) objects they may have also fitted the field. The surface density of a tidally truncated cluster should diminish with radius and eventually go to zero at the tidal radius. However, as can be seen in Figure 2.2b, where I reproduce Figure 8 of Adams et al. (2001), their surface density remains almost constant after 5° . This may be an indication of contamination in their sample.

The two points mentioned before are tightly related. Since so far the surveys have not reached the end of the Pleiades spatial distribution, the samples of members we currently have are spatially biased. They are not a sample of the entire cluster population but just of its central region. Thus, estimates from the King's tidal radius will be biased towards

lower values until the physical end of the cluster will be within the survey's coverage. This issue will be addressed once the full sky coverage of *Gaia*'s data will be available.



(a) Surface density according to [Converse & Stahler \(2010\)](#). The dots and the solid line ([2001](#)). The line shows the fitted King's profile. Reproduced from Figure 8 of [Adams et al. file](#), respectively. Reproduced from Figure 1 ([2001](#)), "The Mass and Structure of the Pleiades of [Converse & Stahler \(2010\)](#)", "The dynamical Star Cluster from 2MASS", AJ, 121. evolution of the Pleiades", MNRAS, 405.

Figure 2.2: Spatial distribution of the Pleiades cluster.

2.4 Velocity Distribution

The three dimensional velocity distribution of the Pleiades has also been studied using its projections. One of them goes along the line of sight, which corresponds to the radial velocity, and another one, which is perpendicular to the previous one, rests in the plane of the sky. This corresponds to the proper motions. Proper motions are angular velocities obtained after measuring the angular displacement that an object shows when measured in at least in two different epochs. Again, measuring the individual stellar positions and its displacement over time is far more easy than measuring radial velocities. These are estimated by means of the Doppler shifting of the absorption lines in the spectre of the object, which means they are observationally intensive. Despite that they are more precise than proper motion measurements (usually on the $1 \text{ km} \cdot \text{s}^{-1}$ regime), they demand large telescopes and longer observing times. For these reasons, historically, the velocity distribution has been studied through the spatial velocity or proper motions distribution.

Probably the first description of the spatial velocity distribution of the Pleiades is that of [Pritchard \(1884\)](#). Using archival data from Knigsberg (1838-1841), Paris(1874) and Oxford (1878-1880) observatories, together with his own *Differential Micrometer* observations, he was able to observe the relative displacements of 40 Pleiades stars. According to him ([Pritchard 1884](#)): *the relative displacements of these distant suns, although not distinctly*

and accurately measurable in numerical extent, appear to vary both in direction and amount; indicating thereby the mutual influence of a group of gravitating bodies.

Later, [Trumpler \(1921\)](#) used, for the first time, proper motion measurements to identify the members of the Pleiades cluster. He classified objects as candidate members according to the distance they show, in the proper motion space, to the mean proper motion of the cluster. This mean was previously calculated by Boss in his *Preliminary General Catalog* ([Trumpler 1921](#)). So far as my historic research went, the Boss work was the first measurement of an statistic of the spatial velocity distribution of the Pleiades.

Later [Titus \(1938\)](#), using Trumpler's data and archival compilations, was able to measure the dispersion of the proper motions distribution. He estimated it to be $0.79 \text{ mas} \cdot \text{yr}^{-1}$. This was probably, the first measurement of the second moment of the spatial velocity distribution. From this value he then derived a total mass of $260 M_{\odot}$.

In recent years [Pinfield et al. \(1998\)](#) used the velocity dispersion to probe that the cluster was in a state near to the virial equilibrium. Later, [Loktin \(2006\)](#) used the projected radial and tangential velocity components of the spatial velocity distribution of 340 members to claim the absence of evidence for rotation, expansion or compression of the cluster. Also, he also found no evidence to support mass segregation.

Concerning the radial velocities, the first record for the Pleiades correspond to [Adams \(1904\)](#). He measured the radial velocities of six of the most brightest stars. In modern years, the latest compilation of radial velocities from the literature is that made by [Galli et al. \(2017\)](#). This list contains measurements for 394 objects. The distribution of these radial velocities is almost gaussian with a centre at $5.6 \text{ km} \cdot \text{s}^{-1}$.

As I will explain later in this Chapter, the complete velocity distribution is a key ingredient in the understanding of the cluster dynamics. Although, spatial and radial velocities are useful projections of the complete distribution, the dynamical analysis of the cluster demands the complete velocity distribution. In [Galli et al. \(2017\)](#) we provide a list of 64 cluster members with full spatial velocities. In Figure 2.3 I show the distribution of the three projection of these spatial velocities.

2.5 Luminosity Distribution

The study of the distribution of luminosities in the Pleiades started few years later than those of the positions and proper motions. The first record I found on the luminosity distribution is the one of [Trumpler \(1921\)](#) (see Fig. 2.4). He computed the number of stars in each magnitude bin for his two samples of candidate members, those comprising the objects within the central 1° , and those within 1° and 3° , referred as Tables I and II, respectively. The completeness of the inner sample was estimated at 14.5 magnitudes whereas that of the outer one at 9.0 magnitudes, both for the visual band. He observed that the luminosity distributions of these two samples were not alike. The inner sample

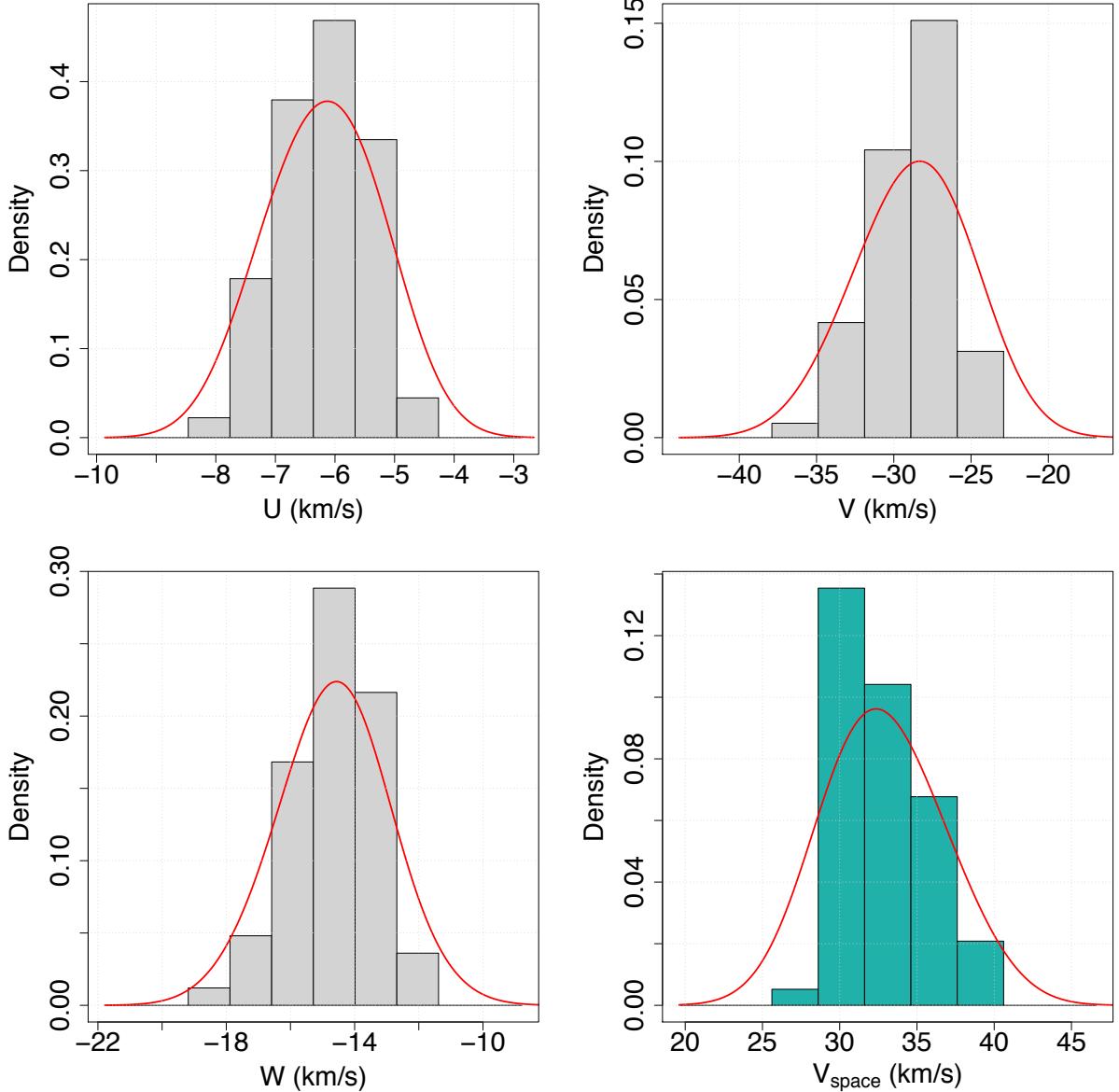


Figure 2.3: Histogram and kernel density estimation (red line) of the components (grey) and modulus (green) of the spatial velocity distribution of 64 candidate members of [Galli et al. \(2017\)](#) with radial velocities and parallaxes. Reproduced from Figure 11 of [Galli et al. \(2017\)](#), "A revised moving cluster distance to the Pleiades open cluster", A&A, 598.

is brighter than the outer one. He also observed that the luminosity distribution is not smooth and shows a local minimum at 9 – 10 magnitudes, then an abrupt rise. Both effects in the two samples.

Later, [Johnson & Mitchell \(1958\)](#) obtained the luminosity distribution using a sample of 289 candidate members. They assessed membership solely on photometry. Their luminosity distribution is shown in Fig. 2.5

Later, [Limber \(1962\)](#) compared the luminosity functions derived from the data of [Trumpler \(1921\)](#), [Hertzsprung \(1947\)](#), and [Johnson & Mitchell \(1958\)](#), with those given by the solar neighbourhood. He noted that the differences between the Pleiades and

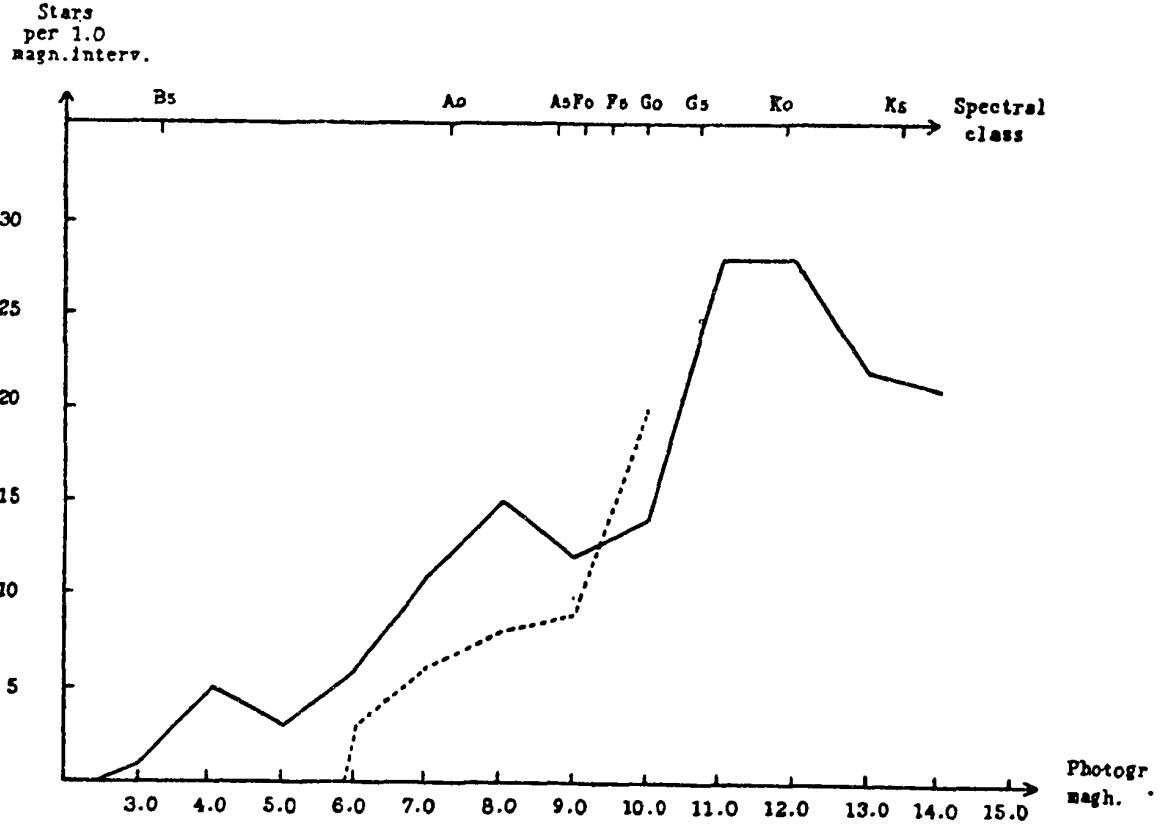


Figure 2.4: Luminosity distribution according to [Trumpler \(1921\)](#). The solid and dashed lines correspond to objects within 1° and within 1° and 3° from the centre. Reproduced from Figure 2 of [Trumpler \(1921\)](#), *The physical members of the Pleiades group*, Lick Observatory Bulletin, 10.

solar neighbourhood start to happen at absolute magnitude 5.5. These distributions were complete until visual magnitudes of 8.5 and 9.5 mag.

In recent years, the luminosity distribution has been described in the works of [Lodieu et al. \(2012\)](#) and [Bouy et al. \(2015\)](#). [Lodieu et al. \(2012\)](#) using the UKIDSS DR9 survey for galactic clusters and a probabilistic members selection method (see discussion in Chapter 1) based on proper motions, and proper motions and photometry, found 8797 and 1147 candidate members, respectively. However, they do not provide the contamination rate in their analysis. Using both lists they provide their luminosity distributions in the Z band, which I show in Fig. 2.6.

In [Bouy et al. \(2015\)](#), we estimated the present day system luminosity distribution of 1378 candidate members contained within the central 3° region (with the centre at $RA = 03 : 46 : 48$ and $Dec = 24 : 10 : 17$ J2000.0). It is called systemic because it has not been corrected for unresolved systems. An unresolved system is a group of stars (e.g. binaries) that due to its compactness appears as a single object. This distribution was computed for the K_s band and is sensitive up to $K_s \sim 20$ mag and complete until $K_s \sim 17$ mag. This luminosity distribution is reproduced in Fig. 2.7

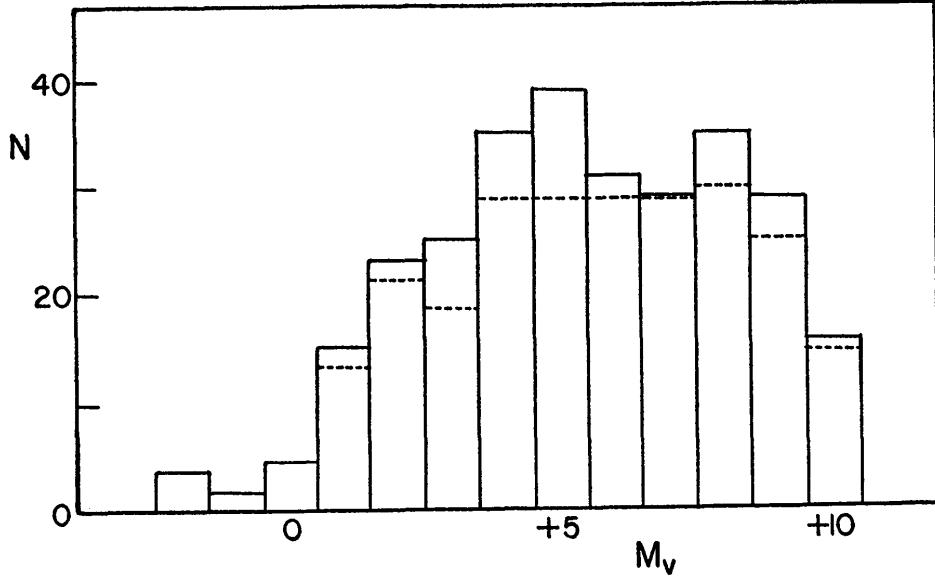


Figure 2.5: Luminosity distribution in the visual band according to [Johnson & Mitchell \(1958\)](#). The dotted line represent the counts of main sequence stars only. Reproduced from Figure 3 of [Johnson & Mitchell \(1958\)](#), "The Color-Magnitude Diagram of the Pleiades Cluster. II.", ApJ, 128.

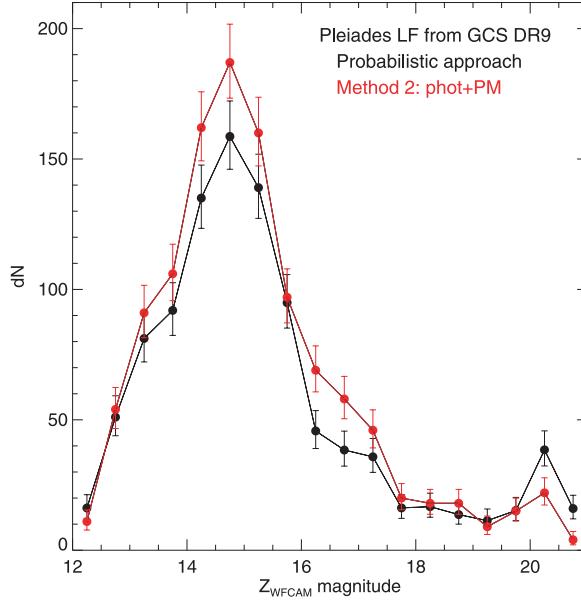


Figure 2.6: Luminosity distribution in the Z band according to [Lodieu et al. \(2012\)](#). The red and black lines correspond to the two different probabilistic methods. Reproduced from Figure 9 of [Lodieu et al. \(2012\)](#), *Astrometric and photometric initial mass functions from the UKIDSS Galactic Clusters Survey - I. The Pleiades*, MNRAS, 422.

2.6 Mass Distribution

In astrophysics, the mass distribution is a cornerstone in the understanding of the star formation process and the later evolution of stellar systems. Although the temporal

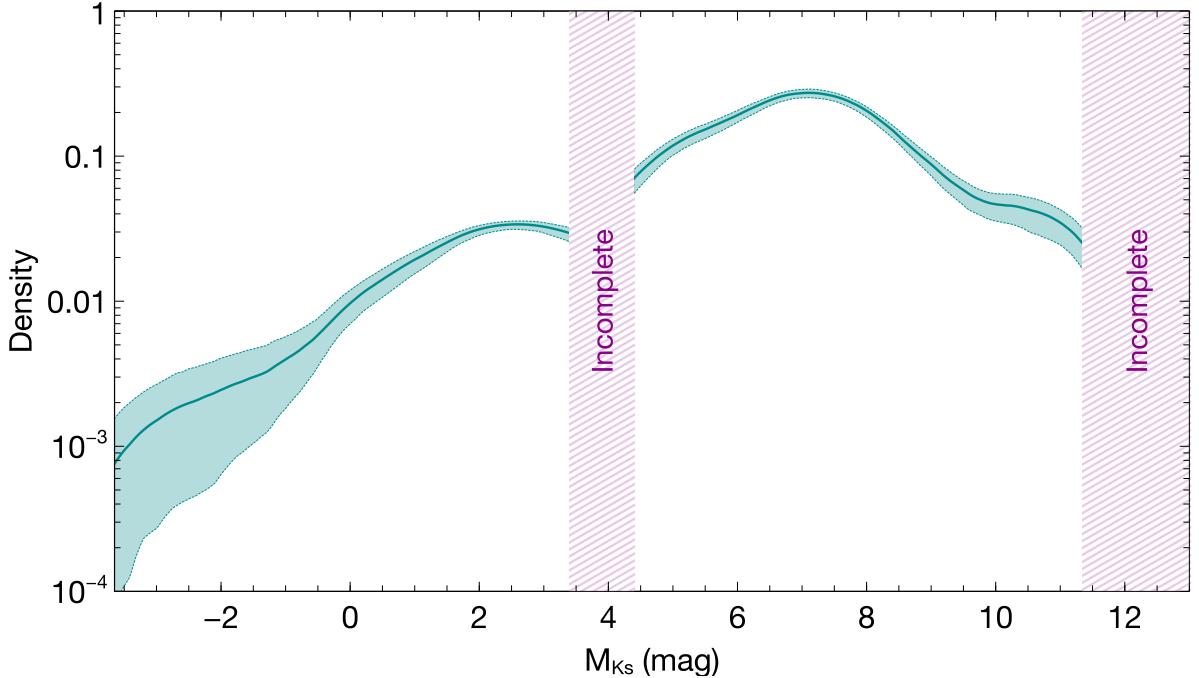


Figure 2.7: Luminosity distribution in the K_s band according to [Bouy et al. \(2015\)](#). The incompleteness regions are shaded. Reproduced from Figure 8 of [Bouy et al. \(2015\)](#), *The Seven Sisters DANCE. I. Empirical isochrones, luminosity, and mass functions of the Pleiades cluster*, A&A, 577.

evolution of these systems is mainly dominated by the gravitational potential, the initial conditions and an ongoing star formation process, if any, can contribute to the shape of the mass distribution. This last contains the fingerprints of past events in the history of the cluster and plays a key role in its future evolution. Indeed, the mass distribution is essential in one of modern astrophysics' objectives: the determination of the role played by the initial conditions or the environment, in the temporal evolution of the stellar systems. Furthermore, the initial mass distribution is a key element in the study of galactic and extragalactic populations.

The mass distribution of the Pleiades has been largely studied. The first work on the mass distribution is that of [Limber \(1962\)](#). Although he did not show any graphical or tabular representation of it, he gave the luminosity distribution and the mass-luminosity ratio. From these the mass distributions can be derived. Instead, he used them to obtain the total mass of the cluster ($760 M_\odot$, see next Section).

Most probably, the first work to present the mass distribution derived from luminosity distributions and a mass-luminosity relation from theoretical models was that of [Hambly & Jameson \(1991\)](#). Using R and I observations from the *United Kingdom Schmidt Telescope Unit* together with the mass-luminosity relation from theoretical isochrone models of Padova group, he was able to transform his luminosity distribution into a mass distribution. In Fig. 2.8, I reproduce his results.

From the year 2000 till date several studies have been published in which the subject

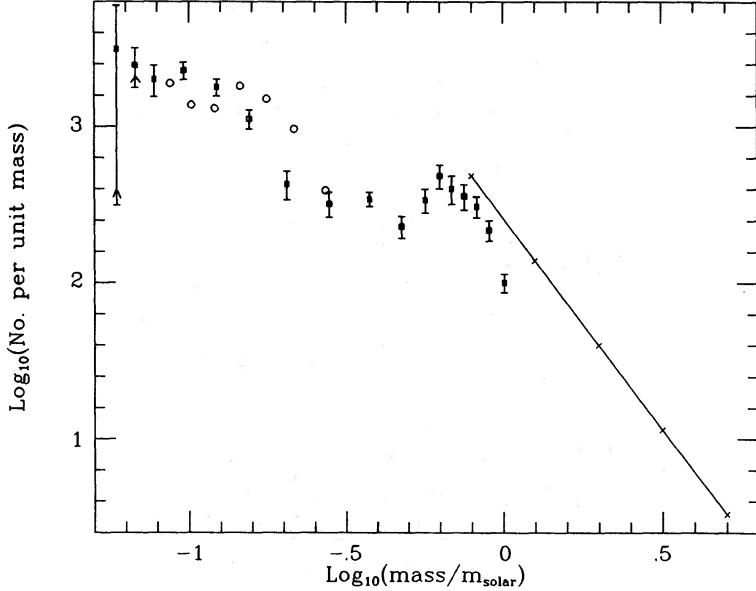


Figure 2.8: Mass distribution of Hambly & Jameson (1991) derived from luminosity distribution and theoretical isochrone models. The open circles result from assuming an older age of 200 *Myr*. The line represent the mass distribution of van Leeuwen (1980). Reproduced from Figure 11 of Hambly & Jameson (1991), "The luminosity and mass functions of the Pleiades - Low-mass stars and brown dwarfs", MNRAS, 249.

of analysis is the Pleiades mass distribution, e.g. Hodgkin & Jameson (2000); Jameson et al. (2002); Moraux et al. (2003, 2004); Lodieu et al. (2007). However, for the sake of simplicity, here I only analyse the two most recent works, those of Lodieu et al. (2012) and Bouy et al. (2015). The mass distributions derived from both these works are shown in Figs. 2.9 and 2.10. Both works obtained first the luminosity distribution, and then transformed it into a mass distributions using theoretical isochrone models.

Lodieu et al. (2012) used a distance of 120.2 *pc*, an age of 120 *Myr*, and the *NEXTGEN* theoretical models of Baraffe et al. (1998) to transform the luminosity into the mass distribution. On the other hand, in Bouy et al. (2015) we use a distance of 136.2 *pc* an age of 120 *Myr* and the *BT-Settl* theoretical isochrone models of Allard (2014).

Both works found contrasting aspects in their discussions. In one hand Lodieu et al. (2012) found that their present day mass distribution agrees with previous studies from the literature, and is also consistent with the system field mass function of Chabrier (2005). Chabrier (2005) found his function by fitting a log-normal function to the visual luminosity distribution of the closest 8 *pc* field objects. On the other hand, in Bouy et al. (2015) we found that although the Chabrier (2005) mass function match that of the Pleiades in the 0.02 – 0.6 M_{\odot} mass range, it predicts to many low-mass stars and brown dwarfs.

The difference between both Pleiades present day mass distributions could arise from the different samples of members, the different theoretical isochrone models, or from both of them. The distance values used in these works can not account for the observed

differences since they introduce only a general shift in the luminosity.

Concerning the differences between the two isochrone models, in Allard et al. (2013) the authors show there are clear differences between the effective temperatures delivered by the *BT-Settl* and the *NEXTGEN* model in the low-mass regime at 5 Gyr.

Concerning the differences between the lists of candidate members, in one hand Lodieu et al. (2012) do not provide (at least explicitly) any estimate of contamination rate of their samples. Furthermore, their membership methodology has some draw-backs (see Sarro et al. 2014) that may have biased their results. Therefore, the agreement that Lodieu et al. (2012) found between their present day mass distribution and the one of Chabrier (2005), which models the field mass distribution, seem to indicate at least the following options. The Pleiades present day mass distribution indeed follows the field mass distribution. Or, the sample of candidate members of Lodieu et al. (2012) is contaminated by the field, thus resembles it.

On the other hand, in Bouy et al. (2015) we estimated a contamination rate of 7%. However, we have no evidence for it to be non-homogeneous. Even if the 7% contaminants were not homogeneous distributed in the mass range, its value is not able to account for the observed discrepancies (30 – 40% in the low-mass regime) between the Chabrier (2005) mass function and our present day mass distribution.

The previous studies show that there is still work to do in the analysis of the Pleiades mass distribution, particularly at the low-mass range where the theoretical models, both of mass function and isochrones, led to discrepancies in the present day mass distribution.

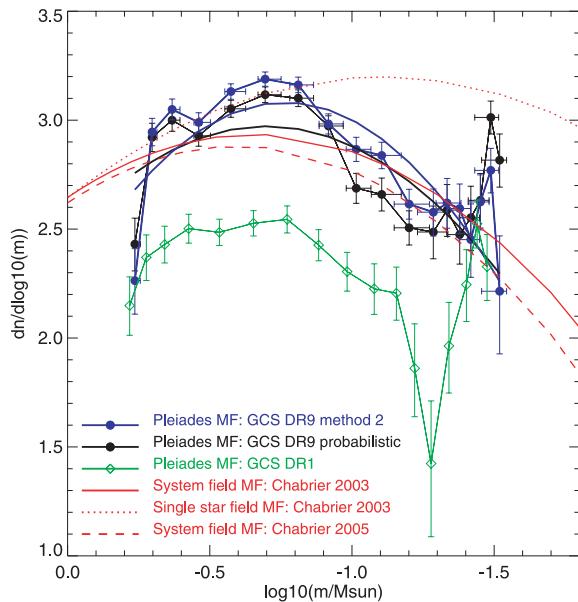


Figure 2.9: Pleiades present day mass distribution from Lodieu et al. (2012). GCS stands for Galactic Cluster Survey. Reproduced from Figure 9 of Lodieu et al. (2012), *Astrometric and photometric initial mass functions from the UKIDSS Galactic Clusters Survey - I. The Pleiades*, MNRAS, 422.

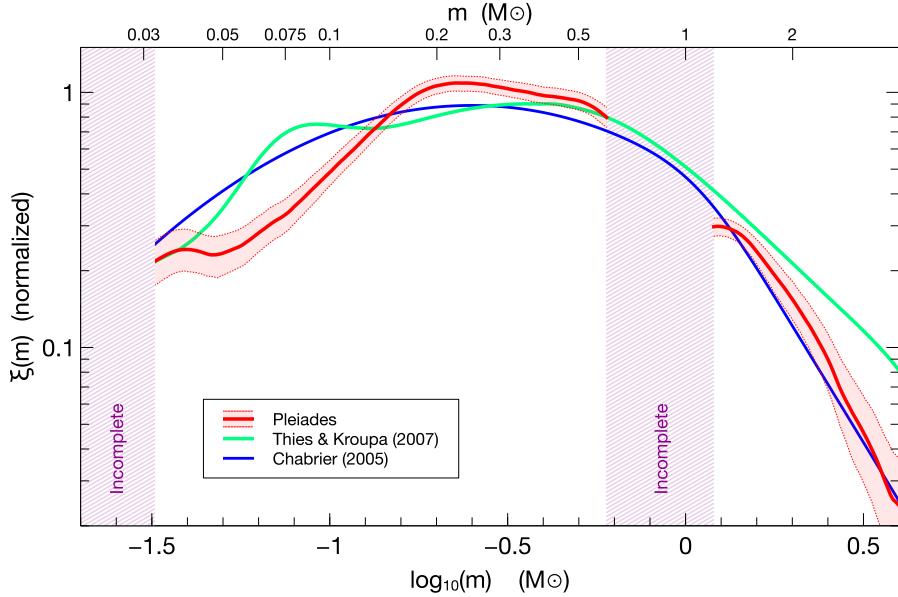


Figure 2.10: Pleiades present day mass distribution from [Bouy et al. \(2015\)](#) (red). IMFs from [Chabrier \(2005\)](#)(blue) and [Thies & Kroupa \(2007\)](#) (green) are also shown. Reproduced from Figure 9 of [Bouy et al. \(2015\)](#), *The Seven Sisters DANCe. I. Empirical isochrones, luminosity, and mass functions of the Pleiades cluster*, A&A, 577.

2.6.1 Total mass of the cluster

Before ending this section I present a (non exhaustive) summary of the studies that provided an estimate of the total mass of the cluster.

The first record I found of the cluster total mass is that of [Titus \(1938\)](#). He estimated a total mass of $260 M_\odot$ assuming virial equilibrium. He also computed $200 M_\odot$ using the Eddington's mass-luminosity relation for objects brighter than 15 mag in the visual band.

The subsequent works continue to report higher masses. [Woolley \(1956\)](#) estimated a total mass of $337 M_\odot$ using a polytrope model fitted to Hertzsprung's catalogue. He then mentions that taking into account Trumpler's data, the total mass should be about $500 M_\odot$.

[Limber \(1962\)](#) computed the total mass in two ways. In the first one he assumed the cluster was virialised and obtained a mass of $900 M_\odot$. Using the luminosity function he estimated the lower limit to the total mass in $760 M_\odot$.

[Jones \(1970\)](#) measured $470 M_\odot$ and $690 M_\odot$ using the luminosity distribution and the virial theorem, respectively.

[van Leeuwen \(1980\)](#) determined a total mass of $2000 M_\odot$ using the virial theorem, a mean individual mass of $2 M_\odot$, and a velocity dispersion of $0.7 km \cdot s^{-1}$ in each spatial direction.

[Lee & Sung \(1995\)](#) measured $700 M_\odot$ using the luminosity distribution and a mass-luminosity relation.

[Pinfield et al. \(1998\)](#) fitting a King's profile to the spatial distribution of the cluster

members obtained $735 M_{\odot}$.

[Adams et al. \(2001\)](#) counting individual masses of candidate members within 5.5° obtained a total mass of $690 M_{\odot}$.

[Converse & Stahler \(2008\)](#) found $820 M_{\odot}$ after adding the individual masses of 1245 candidate members of [Stauffer et al. \(2007\)](#). To obtain these masses they transformed the K and $I - K$ magnitude and colour into masses using the mass-luminosity relation given by the theoretical isochrone models of [Baraffe et al. \(1998\)](#). Later, [Converse & Stahler \(2010\)](#) they redo their analysis and found the total mass to be $870 \pm 35 M_{\odot}$.

2.7 The Pleiades DANCe DR2

The Pleiades DANCe DR2 contains astrometric (stellar positions and proper motions) and photometric ($ugrizYJHK_s$) measurements for 1,972,245 objects. As explained in Chapter 1, the DANCe data set has an heterogenous origin as can be seen from Fig. 2.11. The interested reader can find more the details of this data set and its processing on [Bouy et al. \(2013\)](#). Here I briefly summarise its properties. Table 2.1 contains the basic statistics for the observables, while Table 2.2 does it for the uncertainties. As an example, Fig. 2.12 shows the proper motions uncertainties as a function of the i magnitude.

Table 2.1: Summary of the Pleiades DANCe DR2.

Observable	Min.	1st. Qu.	Median	Mean	3rd. Qu.	Max.	NA's
RA [deg]	51.23	55.40	57.35	57.26	59.01	62.94	0
Dec. [deg]	19.12	22.47	24.32	24.27	25.95	29.69	0
$\mu_{\alpha} [mas \cdot yr^{-1}]$	-99.998	-6.060	-1.645	-1.240	3.401	99.996	0
$\mu_{\delta} [mas \cdot yr^{-1}]$	-99.997	-2.835	2.548	1.976	7.017	99.989	0
u [mag]	13.6	20.4	22.0	21.6	23.3	25.2	1756374
g [mag]	9.4	19.6	22.1	21.1	23.3	25.5	1492564
r [mag]	8.4	17.6	21.3	20.3	22.6	25.1	1222853
i [mag]	7.5	20.0	21.6	21.0	22.7	25.5	820861
z [mag]	11.2	17.9	19.3	18.9	20.2	25.0	697412
Y [mag]	8.3	17.2	18.5	18.1	19.4	24.2	688144
J [mag]	2.8	16.7	17.9	17.5	18.8	23.1	645469
H [mag]	2.0	16.1	17.3	16.9	18.1	20.9	653682
K_s [mag]	1.8	16.0	17.0	16.7	17.7	23.8	561745

2.7.1 Selection of observables

As mentioned earlier, the Pleiades DANCe DR2 contains the positions $R.A., Dec.$, proper motions, $\mu_{R.A.}, \mu_{Dec.}$, and photometric $ugrizYJHK_s$ bands, of almost 2 million sources on the vicinity of the Pleiades clusters. Although these 13 observables carry

KPNO/Mosaic1 **UKIRT/WFCAM** **Subaru/SuprimeCam** **CFHT/CFHT12K**
INT/WFC **CFHT/UH8K** **KPNO/NEWFIRM** **CTIO/MOSAIC2** **CFHT/MegaCam**

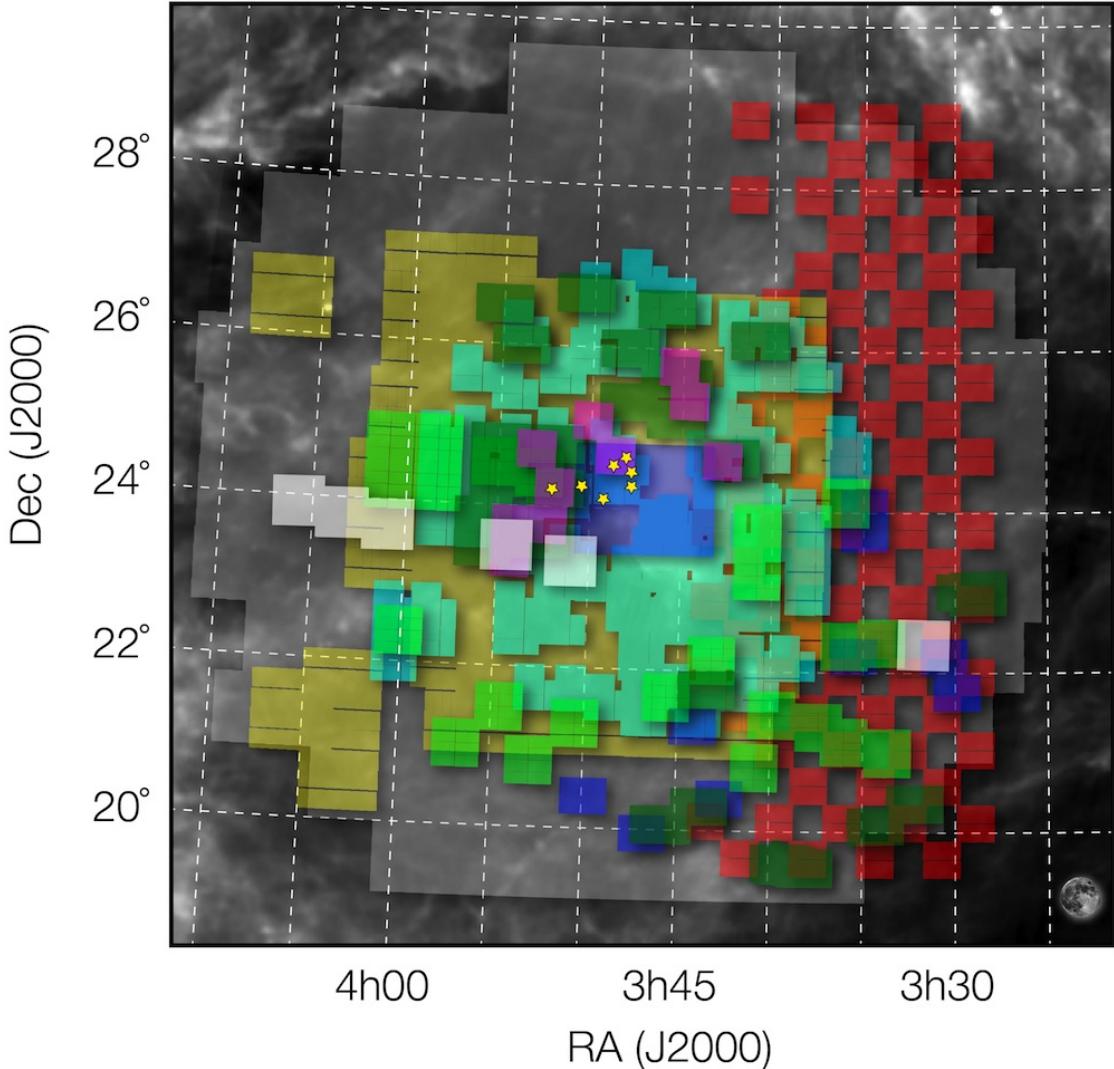


Figure 2.11: Patchy composition of the Pleiades DANCe DR2. The moon shows the scale, and the yellow stars correspond to the central brightest objects of the Pleiades cluster. Reproduced from Figure 1 of [Bouy et al. \(2013\)](#), *Dynamical analysis of nearby clusters. Automated astrometry from the ground: precision proper motions over a wide field*, A&A, 554.

information valuable to discriminate cluster members from field objects, not all of them discriminate in the same amount. [Sarro et al. \(2014\)](#) used important analysis with random forest to select the observables that were the most discriminants. They obtain that the AM, RF-2, and RF-3 reference sets containing the observables $\mu_{R.A.}, \mu_{Dec}$. proper motions, and photometric bands $rizYJHK_s$ were the most discriminants. However, these authors later excluded the r band because most of objects in their training set do not have this band. Since most of the missing values in the r band occur at the faint end, their resulting training set was bised towards the bright end. In a subsequent analysis using roughly the

Table 2.2: Uncertainties of the Pleiades DANCe DR2.

Observable	Min.	1st. Qu.	Median	3rd. Qu.	Max.	Mean
RA [deg]	8.900e-08	9.270e-07	1.933e-06	4.037e-06	2.156e-02	3.173e-06
Dec. [deg]	8.900e-08	9.270e-07	1.932e-06	4.037e-06	2.156e-02	3.173e-06
$\mu_\alpha [mas \cdot yr^{-1}]$	2.01e-01	1.89e+00	4.35e+00	1.00e+01	1.49e+22	3.99e+16
$\mu_\delta [mas \cdot yr^{-1}]$	1.92e-01	1.89e+00	4.35e+00	1.00e+01	4.71e+09	1.42e+04
u [mag]	3.73e-04	8.07e-03	3.06e-02	8.56e-02	2.17e-01	5.48e-02
g [mag]	1.72e-01	1.02e-02	3.90e-02	7.90e-02	1.82e+00	5.34e-02
r [mag]	2.83e-04	1.54e-02	4.88e-02	1.04e-01	1.42e+00	6.34e-02
i [mag]	4.04e-04	9.03e-03	2.73e-02	5.85e-02	2.40e+00	4.37e-02
z [mag]	6.49e-04	5.62e-02	9.16e-02	1.85e-01	3.12e+00	1.34e-01
Y [mag]	3.00e-02	5.21e-02	6.50e-02	1.03e-01	9.01e+00	8.56e-02
J [mag]	1.60e-02	5.24e-02	6.66e-02	1.04e-01	8.89e+00	8.57e-02
H [mag]	1.40e-02	5.28e-02	7.04e-02	1.10e-01	1.00e+01	8.85e-02
K_s [mag]	1.40e-02	5.75e-02	8.17e-02	1.32e-01	3.88e+01	1.04e-01

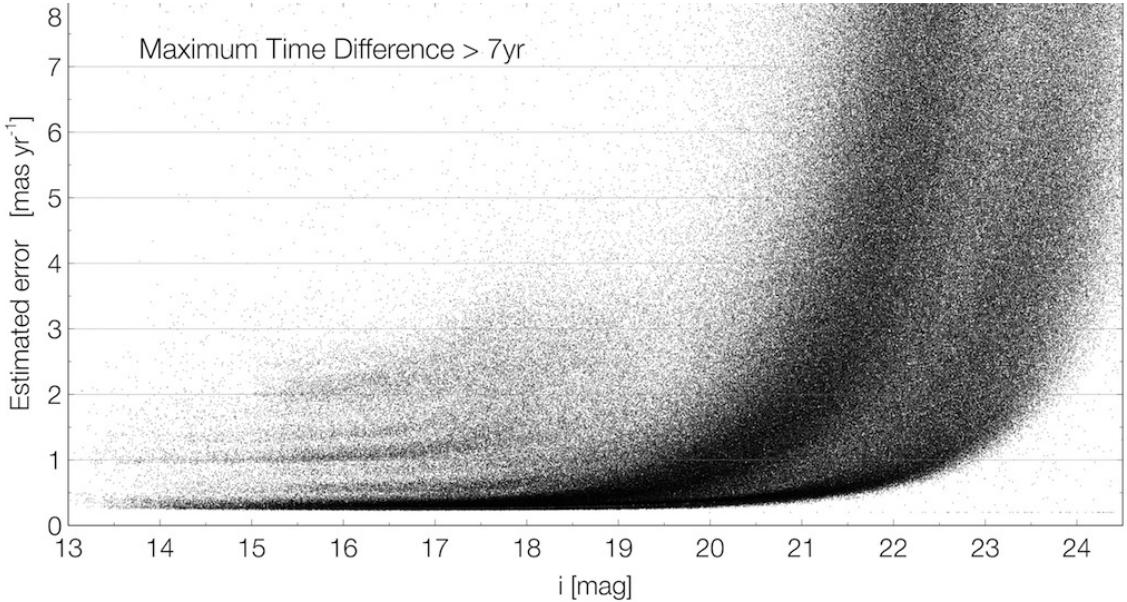


Figure 2.12: Proper motion uncertainty as a function of the photometric magnitude in the i band. Reproduced from Figure 12 of [Bouy et al. \(2013\)](#), *Dynamical analysis of nearby clusters. Automated astrometry from the ground: precision proper motions over a wide field*, A&A, 554.

same methodology, [Bouy et al. \(2015\)](#) worked only the RF-2, which excludes also the z band.

In this work I select the $\mu_\alpha, \mu_\delta, i, Y, J, H, K_s$ observables as my reference set, because these were the ones used [Bouy et al. \(2015\)](#). This selection aims to compare our results with the previous ones of [Sarro et al. \(2014\)](#), and [Bouy et al. \(2015\)](#). However, for the analysis of the spatial distribution we later added the stellar positions (*R.A., Dec.*).

As will be described in Section 3.3.4, the photometry is modelled by parametric series of cubic splines. I choose the *true* colour index $i - K_s$ (in the following *CI*) to be the parameter of these series. This colour allows the most one-to-one dependent-independent variable relation. This one-to-one relation can be seen in Figures 2.13 and 2.14, where I show the colour-magnitude diagram of K vs $i - K_s$ and K vs colours $Y - K_s$, $J - K_s$, $H - K_s$, and $Y - J$, respectively. This one-to-one relation is crucial to avoid degeneracies. Without it, two magnitudes could be described by the same colour index. Therefore a monotonic relation would not be valid.

Thus, our photometric set of observables is made of $i - K_s, Y, J, H$ and K_s .

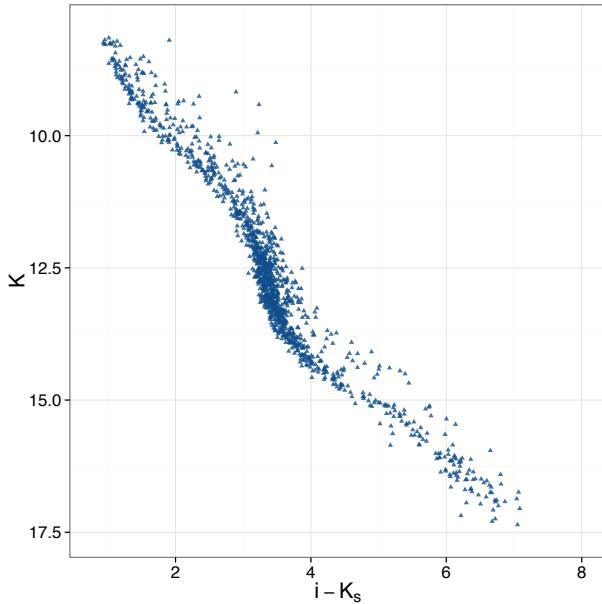


Figure 2.13: K_s vs $i - K_s$ CMD for the Pleiades candidate members of [Bouy et al. \(2015\)](#)

2.7.2 Data preprocessing

Since both photometry and proper motions carry crucial information for the disentanglement of the cluster population, we restrict the data set to only those objects with observed proper motions, and also at least two photometric entries in our photometric set ($i - K_s, Y, J, H, K_s$). These restrictions exclude 22 candidate members of [Bouy et al. \(2015\)](#). Unfortunately, these objects have only one observed value in the photometry. For these particular objects, we compute their marginal proper motion membership probability a posteriori, once the parameters of the model were inferred. The mode and 16 and 84 percentiles of their membership probabilities are listed in Table 2.3. Only four of these 22 objects have membership probabilities below 0.5, which may indicate they are contaminants. Since these membership probabilities were computed using only the kinematic information, these four members could not be discarded as probable candidate members.

Table 2.3: Membership probabilities of the 22 excluded candidate members of [Bouy et al. \(2015\)](#).

ID DANCe	P_{16}	Mode	P_{84}
J035106.55+211604.3	0.751182	0.7774335	0.785560
J035057.42+240630.8	0.792295	0.8090186	0.829541
J034704.76+252249.8	0.701193	0.7319624	0.750745
J034725.80+250832.7	0.789872	0.8169538	0.827954
J034437.44+250815.6	0.762013	0.7773114	0.798125
J035125.88+244738.6	0.883488	0.9007972	0.905823
J034235.64+215029.7	0.838538	0.8662402	0.871439
J034516.66+243432.1	0.862284	0.8666611	0.881004
J034926.12+235714.8	0.852537	0.8606685	0.866286
J034920.60+244635.9	0.923319	0.9270399	0.935532
J035300.63+233252.3	0.762996	0.7747163	0.781901
J034606.52+235020.2	0.928688	0.9333306	0.940772
J035040.89+245657.7	0.509435	0.5215143	0.530379
J034845.33+233124.8	0.260551	0.2650812	0.275513
J034713.67+234953.3	0.814689	0.8489593	0.855902
J034546.48+234743.0	0.897035	0.9098442	0.912059
J034548.95+235110.2	0.933892	0.9376429	0.945558
J035202.26+242148.1	0.248011	0.2649874	0.305949
J035313.22+235540.8	0.518388	0.5425345	0.553376
J034425.60+244052.5	0.581242	0.5902633	0.602096
J035518.38+245637.2	0.198074	0.2087989	0.252831
J035418.93+252944.0	0.366009	0.3760922	0.386213

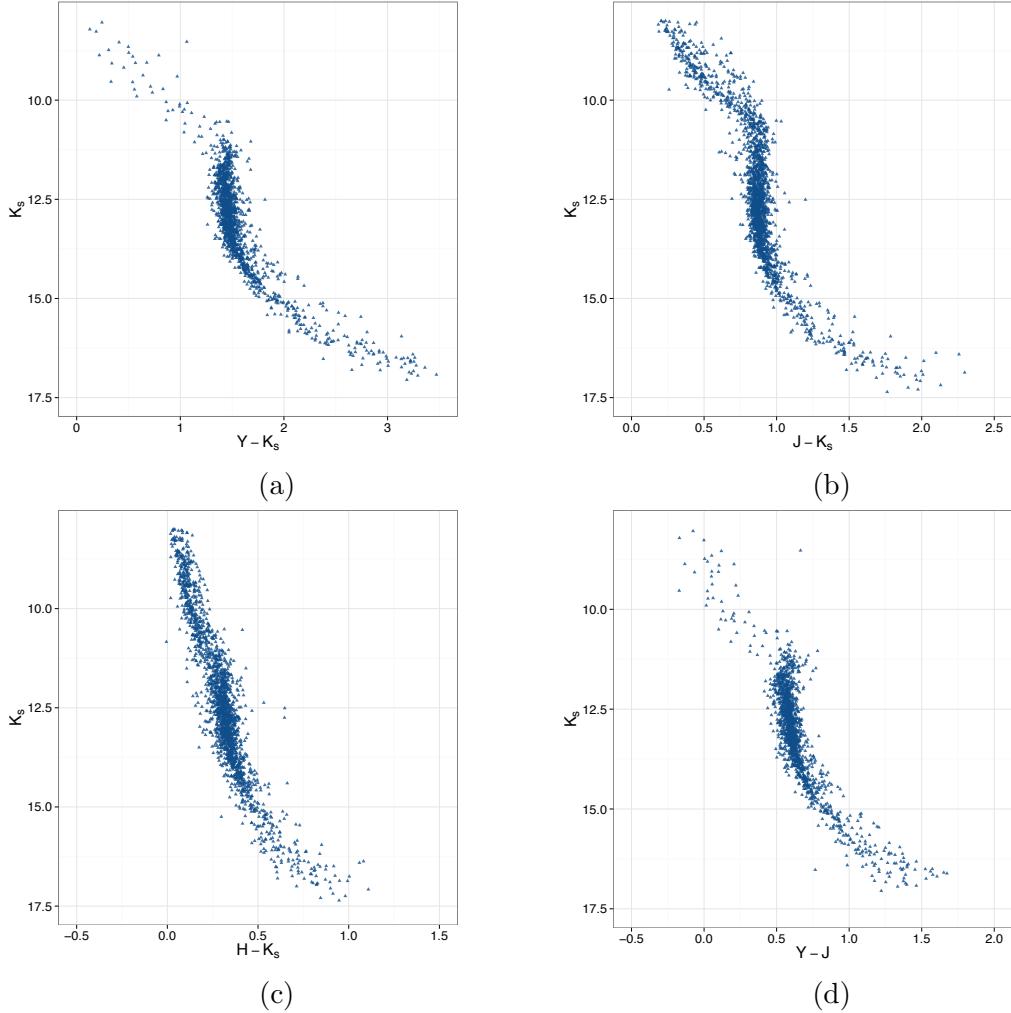


Figure 2.14: CMD for the Pleiades candidate members of [Bouy et al. \(2015\)](#). The magnitude K_s is shown versus the colour indices: $Y - K_s$ (a), $J - K_s$ (b), $H - K_s$ (c), and $Y - J$ (d).

Furthermore, we restrict the lower limit of the CI to the value of the brightest cluster member, $CI = 0.8$. We do not expect to find new bluer members in the bright part of the CMDs. Also, we set the upper limit of the CI at one magnitude above the colour index of the reddest known cluster member, $CI = 8$, thus allowing for new discoveries. Due to the sensitivity limits of the DR2 survey in i and K_s bands, ($i \approx 23$ mag and $K_s \approx 18$ mag, see Appendix A of [Bouy et al. 2015](#)), objects with $CI > 8$ have K_s magnitudes ≥ 16 mag. This combination of CI and K_s magnitude is incompatible with the cluster sequence. Thus, we discard a priori these 695 objects as cluster members. With the previous restrictions to the DANCe DR2, it was reduced to 1 424 893 objects.

Our computational constraints and the costly computations of our methodology (see Sect. 3.2), prevented its application to the entire data set. However, since the precision of our methodology, as that of any statistical analysis, increases with the number of independent observations, we find that a size of 10^5 source for our data is a reasonable compromise. Although a smaller data set produces faster results, it also renders a less

precise and potentially more biased model of the field (in the area around the cluster) and therefore, a more contaminated model of the cluster. Thus, our data set was restricted to the 10^5 objects with highest membership probabilities according to [Bouy et al. \(2015\)](#). In the following I refer to this data set as the restricted DR2 or RDR2. The majority of these objects ($\approx 98\%$) belong to the field with cluster membership probabilities about zero, according to [Sarro et al. \(2014\)](#); [Bouy et al. \(2015\)](#). Thus, under the assumption that the membership probabilities given by these authors are correct, the probability of leaving out a cluster member is negligible. For the remaining of the objects in the Pleiades DANCe DR2, we assign membership probabilities *a posteriori*, once the cluster model is constructed.

Chapter 3

Bayesian formalism

This chapter provides a general introduction to probability theory and its application to parametric inference. The objective of this work is to infer the probability distributions of the cluster properties (e.g. luminosity and velocity). Bayes' theorem provides the proper probabilistic framework for the inference of the parameters governing these distributions. The Bayesian framework demands, though, the setting up of prior beliefs about the parameters values. Thus, later in this chapter, I describe the reason why the Bayesian Hierarchical Models are the least subjective approach to settle down the priors. Once the posterior distribution of the parameters in the model has been analytically described, I proceed to describe the MCMC techniques and the particular one I use to sample the posterior distribution.

In the Sections ahead I also provide the details on the assumptions I make to model the data (i.e. the likelihood), and to choose the parameters of the prior distributions. The two final sections focus on the practical issues related to the sampling of the posterior distributions, and the description of the codes I developed.

Partial results of the work presented here have been submitted to the Journal A&A as [Olivares et al. \(2017\)](#). Thus, in the following, whenever I use the pronoun *we*, it refers to the authors of [Olivares et al. \(2017\)](#). However, I use the pronoun I to emphasize that the particular idea or task was done by me.

3.1 Introduction to probability theory.

Uncertainty and probability are closely entangled. Every measurement has an associated uncertainty, otherwise is not a complete measurement ¹. The term uncertainty must not be confused with the term error, which refers to the difference between the measured value of the quantity and its *true* value² ([JCGM 2008](#)). It is commonly agreed that the uncertainty of a measurement can be expressed in a probabilistic basis ([JCGM 2008](#)). It

¹Upper and lower limits are examples of incomplete measurements.

²The true value is that which ideally results when the uncertainty tends to zero.

means that whenever we measure a quantity, lets say a , the distribution of the repeated measurements of a , follows a probability distribution function (pdf, throughout the text I refer to it as a probability distribution or simply a distribution), $p(a)$. This $p(a)$, as any other (statistical) pdf, satisfies the following properties:

Property 1 It has units, those of the inverse of a .

Property 2 $p(a) \geq 0$.

Property 3 $1 = \int_a p(a)da$.

These properties hold regardless of the dimension of a . It means that the joint probability distribution of the uncertainties of all measured quantities of an object is also a probability distribution. Furthermore, they also hold for conditional probability distributions. Let's imagine that we measure the positions, projected in the plane of sky (the plane perpendicular to the line of sight), of one star. The measurements are conditioned on the magnitude, or brightness, of the object we measure. If the object is too bright, like the sun, it saturates the detector and that renders the measurement useless. On the other hand, if the object is too faint, there are not enough photons to trigger a detection. So, the stellar positions in the sky, which we can call a and b because they are two dimensions, follow a conditioned probability distribution, with the magnitude, c , the conditional variable. Therefore, $p(a, b|c)$ must also satisfy:

- It has units of $a^{-1}b^{-1}$.
- $p(a, b|c) \geq 0$.
- $1 = \int_a \int_b p(a, b|c)da \cdot db$.

The link between joint and conditioned probabilities is given by the following symmetric definition:

$$\begin{aligned} p(a, b) &= p(a|b) \cdot p(b). \\ p(a, b) &= p(b|a) \cdot p(a), \end{aligned} \tag{3.1}$$

which can be further conditioned on c to obtain:

$$\begin{aligned} p(a, b|c) &= p(a|b, c) \cdot p(b|c), \\ p(a, b|c) &= p(b|a, c) \cdot p(a|c). \end{aligned} \tag{3.2}$$

If the joint probability of a and b can be factorised, this is

$$p(a, b) = p(a) \cdot p(b), \tag{3.3}$$

then a and b are say to be *independent*. An alternative option is to say that a and b are *independent* if the conditional probability of a on b is $p(a|b) = p(a)$.

Property 3 establishes that the amount³ of probability distribution $p(a)$ spread over the volume of the support of a adds to one. This Property allows the *marginalisation* of any *nuisance* variable. Let's imagine that someone gives us the probability distribution $p(a, b)$, where a and b are the positions of some star in the plane of the sky. If we are interested in, let's say, the mean value of a , we first must get rid of b . For it, we *marginalise* it out in the following way,

$$p(a) = \int_b p(a, b) \cdot db. \quad (3.4)$$

Then we compute the *expected value* of a , $E(a)$, which is identified with the mean of a once we have drawn many realisations from its probability distribution. To compute it, we add all the possible values of a weighted by their probability. This is,

$$E(a) = \int_a a \cdot p(a) \cdot da. \quad (3.5)$$

Once again, these last two equations (3.4 and 3.5) hold if the distributions are conditioned on any other measurement. For example, the magnitude of the object, as we did before.

It is important to recall that the term measurement, and its unavoidable uncertainty, refer not just to directly measured quantities, like the photons (counts) and pixels in a CCD, but also to indirect measurements. Stellar magnitudes and positions in the sky, for example, are indirect measurements derived from the direct measurement of photons, pixels and telescope arrangements. This generalisation also applies to the measurement of parameters in any physical or statistical model, like the one I will describe in the following Sections.

3.1.1 Bayes theorem

The definition of conditioned probability (Eq. 3.2) leads to Bayes' theorem:

$$p(a|b, c) = \frac{p(b|a, c) \cdot p(a|c)}{p(b|c)}. \quad (3.6)$$

³Which could be infinite, like in Dirac's delta.

Integrating on a we find that,

$$\begin{aligned} p(b|c) \cdot \int_a p(a|b, c) \cdot da &= \int_a p(b|a, c) \cdot p(a|c) \cdot da \\ Z \equiv p(b|c) &= \int_a p(b|a, c) \cdot p(a|c) \cdot da. \end{aligned} \quad (3.7)$$

In this last equation Z refers to what is known as the *evidence*. This Eq. illustrates that $p(b|c)$ is a normalisation constant which can be evaluated once $p(b|a, c)$ and $p(a|c)$ are known. This turns out to be very useful, since as we will see on the following Sections, it tells us that to obtain $p(a|b, c)$ we need just a sample of it, that can be later normalised to obtain a proper probability distribution.

3.1.1.1 Models and parametric inference

In a broad sense, models are representation or abstraction of the *a priori* knowledge some one has about something. Sometimes this knowledge is also shared by others. Models are everywhere in our daily life: from the words we speak every day, to the evolution of the species and the general relativity; from a kid's drawing to cosmological models. In science, however, we restrict the concept of model to a mathematical representation of the relations (the *a priori* knowledge) among the entities that the model attempts to describe. These entities include the observables (i.e. the data), and the parameters, if the model is parametric. In this later case, the observables can be reproduced by means of the parameters. Thus, a parametric model can be thought of as a function which relates parameters and observables in a mathematical way.

The act of finding the parameters that reproduce the observables is called parametric inference. This last can focus either on the entire probability distribution of the parameters, or just on some statistics about it (e.g. the maximum a posteriori, also known as MAP, or the mean and variance).

The proper way to obtain the entire probability distribution of the parameters in a model, given the data, is through Bayes' theorem. Thus it is called Bayesian inference. Another example of parametric inference is the maximum likelihood approach. Where the likelihood, which is seen as a function of the parameters, is maximised. Despite that it properly obtains the parameters that reproduce the data, it does not return their probability distribution. Formally, the likelihood is a probability distribution on the data, and just a function on the parameters. Thus for us to obtain the probability of the parameters we must multiply this likelihood times the priors and normalise it. This is what Bayes' theorem does.

In this context, Bayes' theorem is:

$$p(\theta|\mathbf{D}, \mathcal{M}) = \frac{p(\mathbf{D}|\theta, \mathcal{M}) \cdot p(\theta|\mathcal{M})}{p(\mathbf{D}|\mathcal{M})}. \quad (3.8)$$

where θ , \mathbf{D} and \mathcal{M} correspond, respectively, to the parameters in the model, the data which the model tries to describe, and the the model itself.

The term on the left is called the posterior probability distribution of the parameters, θ given the data \mathbf{D} and the model \mathcal{M} . In the right side, the two terms in the numerator are called *likelihood*, $p(\mathbf{D}|\theta, \mathcal{M})$ and *prior*, $p(\theta|\mathcal{M})$. As mentioned earlier, the denominator, $p(\mathbf{D}|\mathcal{M})$, is called the *evidence*.

Formally, the likelihood and the prior are probability distributions on the data \mathbf{D} and the parameters θ , respectively. However, for the posterior to be a probability distribution on the parameters, it only suffices that the product of the likelihood times the prior does not vanish everywhere or be negative anywhere⁴. If these are not probability distributions, they are called *improper* priors or *improper* likelihoods. In the extreme case that their product vanishes everywhere, which may be the case if the prior is terribly specified or if the likelihood does not take proper account of extreme data, the posterior will not be a probability distribution due to a division by zero. Nevertheless, it makes no sense to try to estimate the parameters of a model with zero evidence.

As mentioned before, the likelihood is the probability distribution of the data given the parameters, regardless the size of the data set. Almost always the data set is a collection of measurements of several objects, but it could also be made of just one object. Thus, to obtain the likelihood of the data based on that of the individual objects, some assumption must be made.

When measuring the properties of individual objects in a collection of them, almost always it is assumed that measurements, and therefore their probability distributions, are independent within objects of the collection. For example, if we were to measure the weight of a group of persons, we usually assume that the pdf of the weight of one person, is independent of that of another person. This assumption, however, does not always holds. Imagine for example, that we were conducting a statistical analysis on the length of the objects used to define our unit of length. In this case, the statistical distribution of the length of any object depends on the unit of measurement, which in turn is depends on the distribution of measured lengths of the rest of the objects. Therefore, the individual probability distributions of the length of the objects are not independent within them⁵.

If we assume that the probability distribution of the measured value, d , of a collection of N objects are independent within them, then

$$p(\mathbf{D}) = \prod_{n=1}^N p_n(d_n), \quad n = \{1, 2, \dots, N\} \quad (3.9)$$

with p_n explicitly stating that the individual probability distributions are distinct.

⁴See Property 2. Although negative probabilities may have sense in quantum mechanics. See for example Dirac (1942)

⁵This happens for example in satellite surveys for which their system of reference is defined by their own measurements. To avoid this issue their reference systems are anchored on independent measurements.

Thus, the likelihood of the data, $p(\mathbf{D}|\theta, M)$ can be expressed as,

$$p(\mathbf{D}|\theta, M) = \prod_{n=1}^N p_n(d_n|\theta, M), \quad n = \{1, 2, \dots, N\}. \quad (3.10)$$

The term $p_n(d_n|\theta, M)$ is the likelihood of datum d_n . This is also called the *generative* model, since it contains the necessary information to generate the data⁶.

To summarise, Bayes' theorem can be interpreted as the probabilistic way to update knowledge. To me, it embodies the process of knowledge improvement once we recognise that knowledge is uncertain, which in my perspective, is always uncertain. Even when its uncertainty is negligible under the current evidence that supports it. Bayes' theorem helps us to update our prior knowledge once we multiply it by the likelihood of the data. Then, the posterior probabilities, became our new knowledge.

Furthermore, the Bayes' theorem also provides the objective way to compare two models or hypothesis, and update the a priori knowledge used to construct them. This is called model selection, which I briefly explain in the next section.

3.1.2 Model Selection

Whenever we have a data set and two or more models that attempt to describe this data, the most straightforward thing to do is to compare these models. Almost always, we want to select the *best* model. Obviously the term *best* depends on the objective of research. For example, let's imagine that our data set consists of the positions of an object as function of time. If we were interested in reproducing exactly the same points in the data set, the *best* model will be a polynomial with degree equal to the number of points. This polynomial will pass through all these points. However, once we recognise the unavoidable uncertainty of the data, we realise that an exact representation of the data may be of no use since it fits also the noise of the data.

In general, we are interested in the predictive capabilities of a model, its ability to predict future observations rather than to replicate the ones we currently have. Thus, an exact representation of the observed data (an over-fitted model as in the previous example), will poorly describe any new data set. In this sense, an over-fitted model *memorises* the data rather than *learns* from it.

A model that *learns* from the data is that which recovers the *true* underlying relation embedded in the data. This *true* underlying relation is the one that produces the *true* data. The observed data results once the uncertainty process is added.

Nevertheless, we still need to select among different learning models.

We can draw some help from the commonly known Ockham's razor or principle⁷. It

⁶It may also contain the uncertainty process that transform the *true* data into the noisy observations.

⁷The origin of this motto and its exact phrasing is beyond the scope of this work. I just mention that paradoxically, an ancient formulation is attributed to Ptolomey: "We consider it a good principle to

says:

Among competing hypotheses, the one with the fewest assumptions should be selected.

Here, hypotheses can be identified with models. Thus, this principle tells us we should choose the model that makes the fewest assumptions. I classify the assumptions of a model in two groups: fixed and free ones. The fixed assumptions belong to what I previously described as the *a priori* knowledge used to construct the model. These may render the model more interpretable in the physical or statistical sense, or even give it coherence within the corpus of a theory. The free assumptions on the other hand, correspond directly to the parameters in the model. They give it flexibility when fitting the data⁸. For example, in the case of a straight line model, the fact that the data is linearly related can be considered as a fixed assumption. The free assumptions correspond to the slope and ordinate at the origin.

When comparing a linear model to a quadratic one in which the constant term has been fixed, we see that they have the same number of free parameters, two, but clearly the second one has an extra fixed assumption. Therefore, choosing the model with fewer free parameters does not necessarily mean choosing the model with the fewest assumptions.

One of the great advantages of the Bayesian methodology is that it incorporates directly Ockham's principle. Suppose that we want to compare two models, M_1 and M_2 , which we assume describe the data set, \mathbf{D} . Each model has prior probabilities, $p(M_k)$ and likelihoods $p(\mathbf{D}|M_k)$ (with $k = 1, 2$). Notice that now, I use Bayes' theorem for models and not for parameters within a model. So, the prior probabilities of the models reflect our beliefs about the fixed assumptions within each model. On the other hand, the likelihood of the data given the model, is related to the parameters (the free assumptions) and their prior probabilities, both within a model. This likelihood of the data given the model corresponds to the *evidence* of the model (Eq. 3.7). This evidence in terms of the model parameters, θ_k , is now

$$p(\mathbf{D}|M_k) = \int_{\theta_k} p(\mathbf{D}|\theta_k, M_k) \cdot p(\theta_k|M_k) \cdot d\theta_k. \quad (3.11)$$

Bayes' theorem applied to models, instead of individual parameters as done before, tells us that

$$p(M_k|\mathbf{D}) = \frac{p(\mathbf{D}|M_k) \cdot p(M_k)}{p(\mathbf{D})}, \quad (3.12)$$

with $k = 1, 2$. Since there are only two models, their prior probabilities are related by $p(M_1) = 1 - p(M_2)$. Therefore,

$$p(M_k|\mathbf{D}) = \frac{p(\mathbf{D}|M_k) \cdot p(M_k)}{p(\mathbf{D}|M_1) \cdot p(M_1) + p(\mathbf{D}|M_2) \cdot p(M_2)}. \quad (3.13)$$

explain the phenomena by the simplest hypothesis possible" (Franklin 2002)

⁸However, they can also introduce degeneracy in the parametric space.

From this last Equation, the ratio of the posterior distributions is:

$$\frac{p(M_1|\mathbf{D})}{p(M_2|\mathbf{D})} = \frac{p(\mathbf{D}|M_1) \cdot p(M_1)}{p(\mathbf{D}|M_2) \cdot p(M_2)}. \quad (3.14)$$

This ratio provides an objective measure of how better the model M_1 is when compared to model M_2 , under the measure provided by the data \mathbf{D} by means of the evidence. When both prior probabilities $p(M_1)$ and $p(M_2)$ are set alike, the ratio of posteriors equals the ratio of likelihoods. This is known as the *Bayes factor* (for a similar derivation and some examples of its application see [Kass & Raftery 1995](#)).

Even when the priors for the models are set alike, the evidences themselves (Eq. 3.11) embody Ockham's principle. The evidence is the integral, in parametric space, of the prior times the likelihood, with the likelihood acting as a weight to the priors. Then, the larger the number of parameters (free assumptions) is, the greater the volume over which the integral must be carried on, and the most spread the prior probability gets. Thus, unless the likelihood increases as well, the evidence is smaller in models with larger number of parameters.

As explained in this section, the paramount importance of Bayes' theorem comes from the fact that it is the proper probabilistic way to update knowledge based on evidence.

3.1.3 Membership probability

In the previous Section we obtained the probability of two competing models M_1 and M_2 given the data \mathbf{D} . In this Section, I describe a similar problem: the probability of two competing models given the likelihood of a single datum, \mathbf{d} . It also can be interpreted as the probability that the datum \mathbf{d} was generated by model M_k . This probability is commonly known as the membership probability of the datum \mathbf{d} to belong to model or class, M_k ($k = 1, 2$).

Bayes' theorem for this particular case is,

$$p(M_k|\mathbf{d}) = \frac{p(\mathbf{d}|M_k) \cdot p(M_k)}{\sum_{k=1}^2 p(\mathbf{d}|M_k) \cdot p(M_k)}, \quad (3.15)$$

where $p(\mathbf{d}|M_k)$ is the likelihood of datum \mathbf{d} and, $p(M_k)$ is the prior probability of model M_k .

3.2 Bayesian hierarchical Models

3.2.1 Generalities

The Bayesian formalism requires the establishment of priors. As mentioned before, priors represent the *a priori* beliefs the user of the model has about the possible values

that parameters of the model can take. This is indeed subjective. This subjectivity is the main source of criticism from the non-bayesian community⁹.

Bayesian hierarchical models, in the following (BHM), are classified within the Empirical Bayes methods. On these methods, the prior distributions are inferred from the data rather than being directly specified, as it is done in common Bayesian methods. Particularly, in BHM the priors are specified by parametric distributions whose parameters are also drawn from a parametric distribution in a hierarchical fashion. For this reason, hierarchical models are also called multilevel models. A fully-BHM is that in which the parameters at the higher hierarchy are drawn from a non-parametric distribution. In a non fully-BHM the settlement of parameters stops at some level. These end-level parameters are called hyper-parameters.

Given their properties, the BHM represent the most objective way to the establishment of prior distributions (Gelman 2006). Regardless of the level of the BHM, for it to be effective, the family, or class, of prior distributions must be carefully chosen. These families must allow the *true* value of the parameter of interest (Morris 1983). If the likelihood is not fully allowed by the prior (for example when the likelihood, as a function of the parameters, has a maximum outside the domain of a truncated prior.), then the inferred posterior could be biased. For this reason, inspecting and updating the prior knowledge is an important step in any Bayesian study.

Despite their theoretical advantages, BHM are difficult to evaluate since they require far more parameters than standard Bayesian methods. Furthermore, their hierarchy (levels) must stop at some point. There are at least two approaches to stop this hierarchy. The first one is to use a non parametric distribution for the parameters at the higher level. This renders, as previously said, a fully-BHM. However, it has the inconvenience that to use a non-parametric distribution we must have certain a priori knowledge about it, which, most of the time is not the case. Another more practical alternative is to give a point estimate, usually the mean or the mode, for the distribution of the parameter at the top of the hierarchy.

Although in BHM the parameters values of the prior distributions are inferred from data, the user of the model has the important task of specifying the family of distributions to be used for the priors. Selecting these families continues to be an active area of research. Common options include conjugate, objective, non-informative, and weakly informative priors.

Conjugate priors are those in which the posterior distribution turns out to be in the same family as the prior one, they are called the conjugate of the likelihood. Objective or default priors are supposed to be the default choice in the absence of any other information. Weakly and non-informative informative priors, as their names indicate, provide intentionally weaker information or no information at all for the prior.

⁹See Gelman (2012) for a discussion on the ethical use of prior information

Objective and non-informative priors are discarded since for the Pleiades (one of the most studied cluster in the literature) there is already sensible information that should be used. Conjugate priors do not necessarily agree with the previous information. Thus, whenever possible, I choose the weakly informative priors. These are the option recommended by [Gelman \(2006\)](#); [Gelman et al. \(2008\)](#); [Huang & Wand \(2013\)](#) and [Chung et al. \(2015\)](#), since they show better computational performance when compared to non-informative priors. Despite the family of prior distribution that we choose, we must always analyse the prior distribution in terms of the posterior, and check if this last make sense ([Gelman 2006](#); [Gelman et al. 2013](#), Chap. 6).

3.2.2 Examples

Since BHM usually need more parameters than standard techniques, its use was restricted until modern and faster computers were widely available. The concept of BHM was already present in the 1960s, however, it was not until the 1970s that they were used to infer parameters of normal distributions and linear models (see [Good 1980](#), for an historical perspective of BHM). In modern days, BHM have a wide range of applications. Just to cite some examples, [Gelman & Hill \(2007\)](#) use them in the social sciences, [Fei-Fei & Perona \(2005\)](#) applies them for vision recognition and, [Diard & Bessiere \(2008\)](#) for robot navigation.

The BHM are also widely applied in astrophysics. Although, originally its use was mainly in the domain of inference of cosmological parameters (see for example the works of [Feeney et al. 2013](#); [March et al. 2014](#); [Anderes et al. 2015](#); [Shariff et al. 2016](#); [Alsing et al. 2017](#)), they were rapidly adopted in other domains. For example, they have been used to study the eccentricity distribution of binary stars [Hogg et al. \(2010b\)](#), the Cepheids ([Barnes et al. 2004](#)) and RR Lyrae distances ([Jefferys et al. 2007](#)), the chemical composition ([Wolfgang & Lopez 2015](#)) and albedos of exoplanets ([Demory 2014](#)), extinction maps ([Sale 2012](#)), stellar parameters ([Shkedy et al. 2007](#)), and the present day mass distribution ([Tapiador et al. 2017](#)).

3.2.3 Graphical representation.

In general, BHM have large number of parameters that, together with the relations among them tend to obscure its understanding. To aid in this understanding, a graphical representation is always useful. Probabilistically Graphical Models (PGM) provide the background to graphically depict the BHM. PGM are graphs that portray the conditional relation among the elements in a probabilistic model. These elements could be constants or stochastic variables that interact by means of conditional relations, which could be deterministic or stochastic.

In PGM, stochastic variables are represented with circles while constants with squares.

If the variable is known, as in the case of the data, it is represented with a filled symbol, otherwise with an empty symbol. Stochastic and deterministic relations are depicted with solid and dashed lines, respectively. If there is no line between two given elements, it indicates that they are assumed to be independent. Variables that repeat together, as in the case of the data, are grouped within a plate. The number of repetitions is indicated in one corner of the plate. For more details on PGM see for example the book of [Koller & Friedman \(2009\)](#).

To exemplify the use of PGM, Figure 3.1 shows the PGM for the BHM that infer of the parameters in a Gaussian Mixture Model. In this model the parameters are the means μ_k , variances σ_k^2 and fractions ϕ of each of the K gaussian distributions in the mixture. Then μ_0 , σ_0^2 , λ , ν and β are the hyper-parameters, while x_i represent the data, and z_i the categorical latent variable that indicates the parent gaussian for datum x_i .

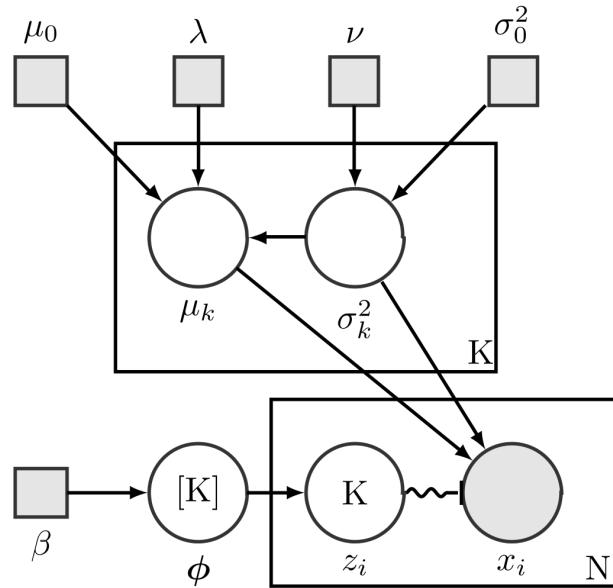


Figure 3.1: PGM representing the parametric inference of a Gaussian Mixture, see text for details. Figure by Benwing /CC BY

3.3 Modelling the data

Creating a model is a complex task. As previously mentioned, a model is the mathematical representation of the knowledge about certain phenomenon. In this work my objective is the statistical modelling of the DANCE data related to Nearby Young Open Clusters (NYOC), particularly the one of the Pleiades (see Section 2.7).

Modelling a data set demands the gathering and arranging of the *a priori* knowledge on the phenomenon. This knowledge comprises that of the data set , the NYOC: the Pleiades open cluster (Chapter 2), the statistical techniques that may help to attain the

objective, and the computational resources at hand. I collect this knowledge from three main sources: the standard references (e.g. articles and books), my colleagues and experts on the field (*knowledge elicitation*), and my self.

The structuring of the a priori knowledge and later creation the model is an iterative and continuous process. In this Section, I describe a snapshot of this process: the state of the model once the article [Olivares et al. \(2017\)](#) was submitted. Later in Section 3.6, I will give a brief description of the chronological development of the model.

In this section I will explain the statistical procedure to deal with one of the crucial aspects of the Pleiades DANCe DR2 data set: the missing values. Also, I will describe how the relevant knowledge about Pleiades cluster population and the empirical one of the field are embedded in the generative model of the data.

3.3.1 Missing values

Missing values refer to a non-measured or non-available (NA) value in the vector of measurements of an object. They can arise due to different statistical or physical processes.

From the physical perspective, they occur due to faint or bright sources that produce counts values outside the dynamical range of the detector. They can also emerge due to detector malfunctions, e.g. electronic failures, or to random effects, e.g. cosmic rays.

From the statistical perspective, important aspects of missing values are, i) the probability distribution of the sources in which they occur, and ii) if they originate due to censoring or truncation.

In the DANCe DR2, missing values occur only in the photometric measurements (see Table 2.1), with the bluer bands being the most affected. As expected, the probability distribution of sources with missing values is not random (uniform). Missing values occur with higher probability at the faintest and brightest ends of the photometric distributions, see Figure 3.2. This is a crucial aspect for the statistical analysis. If missing values were randomly distributed, then, any statistical analysis derived using only objects with completely observed vectors will be unbiased. This will occur because the objects with completely observed vectors will be a random sample of the data set. Since this is not the case in the DANCe DR2, to avoid biases, objects with missing values must be incorporated in the data set and properly treated.

Truncation occurs when the data set does not contain record, nor of the measured value neither of the number of measurements, when the measured quantity lies outside certain limits. On the other hand, censoring happens when the data set contains only partial information about the actual value of the measured quantity. Basically, censoring happens when the value lies outside the upper or lower limits of the measuring instrument, and these limits become the only available record of the measurement. Currently, the DANCe DR2 does not provide upper or lower limits for the censored data. Although these limits could also be inferred from the data, the heterogeneous origin of the DANCE

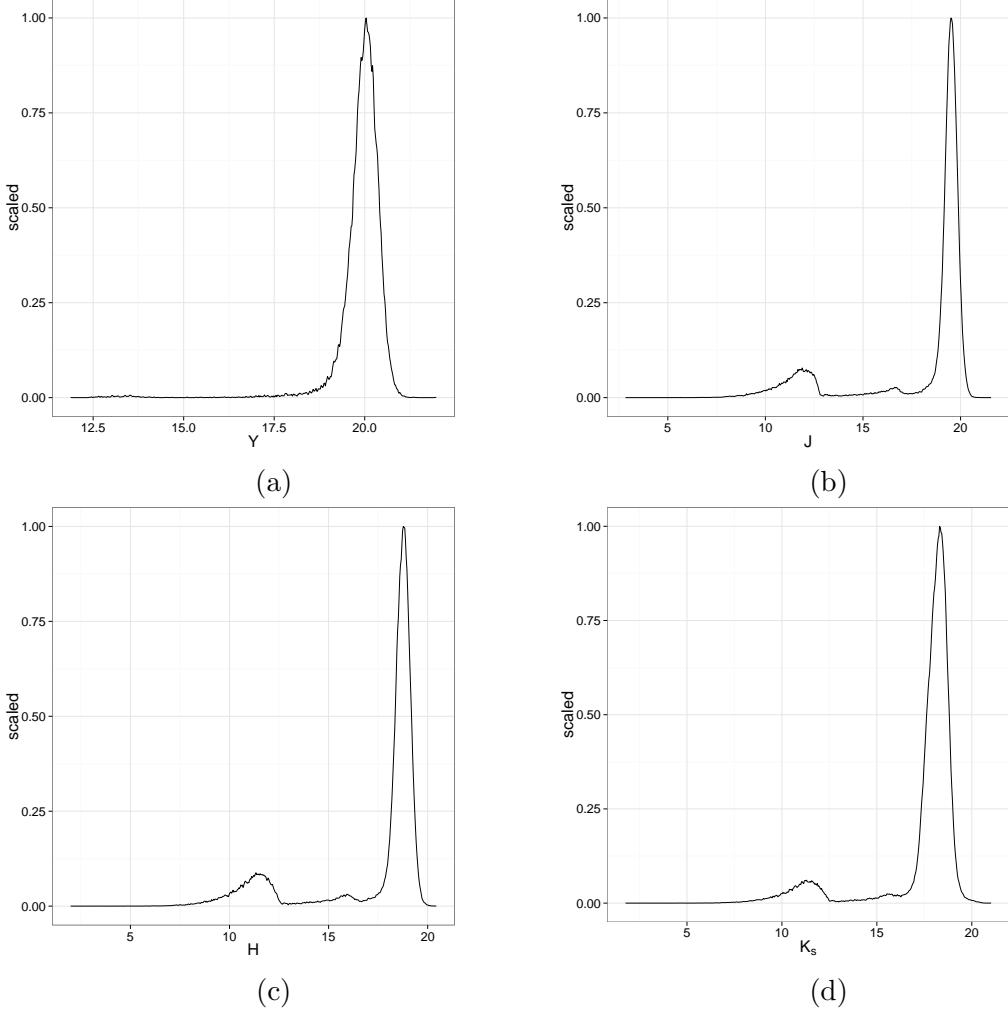


Figure 3.2: Probability distributions of objects with missing values. The vast majority of objects with missing values are located at the faint end of the magnitude distributions.

DR2 prohibits this task. Furthermore, the missing values in the DANCE DR2 data set do not occur only because of censoring. Indeed, cosmic rays, hot pixels, halos of bright stars, diffraction patterns and cross-matching failures are just some examples of potential sources of missing values. The statistical treatment of missing values originating from all these sources lies beyond the scope of this work. However, there is a general way to deal with missing values despite their origin.

In terms of probability, there is no distinction between missing values and parameters. Therefore, we can marginalise missing values as we do with any other nuisance parameter. If the vector of measurements \mathbf{d} has a missing value in entry i , then $x_i = NA$, and $\mathbf{d} = \{x_1, x_2, \dots, NA, \dots, x_n\}$, with n the dimension of \mathbf{d} . Thus, the marginalised likelihood of this datum, given model parameters θ is

$$p(\mathbf{d}|\theta) = \int_{-\infty}^{\infty} p(\{x_1, x_2, \dots, x_i = NA, \dots, d_n\}|\theta) \cdot dx_i. \quad (3.16)$$

Throughout this work, missing values are marginalised in this way.

3.3.2 The generative model

The objective of this work is the statistical study of NYOC, in particular the Pleiades cluster. Since the Pleiades data is mixed with that of the field, to achieve the objective, cluster and field populations must be disentangled. The probabilistic disentanglement of them, demands probabilistic models for each population. These models correspond to the are the field and cluster likelihoods. With these and the prior distributions, I am able to reach the objective: the posterior distributions of the parameters in the cluster model.

In addition, cluster membership probability distributions (Eq. 3.15) can be computed for each object in the data set. Using these membership probability distributions together with a probability threshold, individual objects are separated into cluster and field populations.

The inference process demands a set of N binary integers \mathbf{q} , one q_n for each object. The two possible values of these binary integers represent one of the two mutually exclusive possibilities: the object belongs to the cluster ($q_n = 1$) or to the field population ($q_n = 0$).

Let θ and ϕ be the parameters of the cluster and field models, and p_c and p_f their likelihoods, respectively. Then, the likelihood of the data is,

$$p(\mathbf{D}|\mathbf{q}, \theta, \phi) = \prod_{n=1}^N p_c(\mathbf{d}_n|\theta)^{q_n} \cdot p_f(\mathbf{d}_n|\phi)^{(1-q_n)}. \quad (3.17)$$

The inference of these N binary integers will demand a computing power that is outside the current possibilities. Thus, instead of inferring them I marginalise them.

I marginalise these \mathbf{q} using a prior probability, which is set in terms of a new and unique parameter π . It represent the *prior* probability that an object belongs to the field. Thus, the prior probability of \mathbf{q} is

$$p(\mathbf{q}|\pi) = \prod_{n=1}^N (1 - \pi)^{q_n} \cdot \pi^{(1-q_n)}. \quad (3.18)$$

The marginalisation runs as

$$\begin{aligned} p(\mathbf{D}|\pi, \theta, \phi) &= \int_{\mathbf{q}} p(\mathbf{D}, \mathbf{q}|\pi, \theta, \phi) \cdot d\mathbf{q} \\ &= \int_{\mathbf{q}} p(\mathbf{D}|\mathbf{q}, \pi, \theta, \phi) \cdot p(\mathbf{q}|\pi) \cdot d\mathbf{q} \\ &= \int_{\mathbf{q}} \prod_{n=1}^N p_c(\mathbf{d}_n|\theta)^{q_n} \cdot p_f(\mathbf{d}_n|\phi)^{(1-q_n)} \cdot \prod_{n=1}^N (1 - \pi)^{q_n} \cdot \pi^{(1-q_n)} \cdot d\mathbf{q} \\ &= \int_{\mathbf{q}} \prod_{n=1}^N [(1 - \pi) \cdot p_c(\mathbf{d}_n|\theta)]^{q_n} \cdot [\pi \cdot p_f(\mathbf{d}_n|\phi)]^{(1-q_n)} \cdot d\mathbf{q} \\ &= \prod_{n=1}^N (1 - \pi) \cdot p_c(\mathbf{d}_n|\theta) + \pi \cdot p_f(\mathbf{d}_n|\phi). \end{aligned} \quad (3.19)$$

This last equality is a rather complicated derivation which can be found on Press (1997) and in Hogg et al. (2010a) for p_c and p_f in the exponential family. Also, a general derivation of this expression is given by Jaynes (2003). He obtains it assuming individual unknown probabilities p_n instead of q_n and marginalising over them by the aid of a prior.

Thus, the *generative model* or likelihood of the datum \mathbf{d}_n is

$$p(\mathbf{d}_n|\pi, \boldsymbol{\theta}_c, \boldsymbol{\theta}_f, \mathbf{u}_n) = \pi \cdot p_f(\mathbf{d}_n|\boldsymbol{\theta}_f, \mathbf{u}_n) + (1 - \pi) \cdot p_c(\mathbf{d}_n|\boldsymbol{\theta}_c, \mathbf{u}_n), \quad (3.20)$$

where $\boldsymbol{\theta}_f$ and $\boldsymbol{\theta}_c$ indicate the cluster and field parameters, while \mathbf{u}_n refers to the datum uncertainty. The probabilities $p_f(\mathbf{d}_n|\boldsymbol{\theta}_f, \mathbf{u}_n)$ and $p_c(\mathbf{d}_n|\boldsymbol{\theta}_c, \mathbf{u}_n)$ are the field and cluster models, respectively. These models are explained in detail in the next two sections.

In the following, I assume that the observed quantities, even if they contain missing values, resulted from the convolution¹⁰ of the *true* quantities with a source of uncertainty. This last could be different for each object, thus resulting in different individual uncertainties. It means that the DANCE DR2 data is modelled allowing for its intrinsic heteroscedasticity. Although these uncertainties are not assumed to have the same values, I do assume that they have the same shape. Thus, I assume that uncertainties in our data set belong to the multivariate normal family. This assumption is standard practice and is also supported by the large and heterogeneous origins of the DANCe DR2 data set.

3.3.3 The field population

To model the field population, I assume that the joint seven-dimension probability distribution of the field can be factorised into the probability distributions of proper motion and photometry. Thus, the field likelihood of the proper motion and the photometry are assumed to be independent, at least conditioned on their parameters. Also, I assume that both distributions are described by Gaussian Mixture Models (GMM). The flexibility of GMM to fit a variety of distribution geometries make them a suitable model to describe the field density of the heterogeneous DANCe DR2 data.

A GMM is a probability distribution resulting from the linear combination of M gaussian distributions, $\pi_m \cdot \mathcal{N}(\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m)$, with $m = 1, 2, \dots, M$. Where π_m is the amplitude or fraction of the m th gaussian. These m fractions must add to one. Thus, the probability distribution resulting from a GMM is

$$p_{GMM}(x|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{m=1}^M \pi_m \cdot \mathcal{N}(\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m). \quad (3.21)$$

Notice that the number of gaussians in the mixture is not formally speaking a parameter, but rather it implies a collection of parameters. The number of these parameters increases

¹⁰The addition of two stochastic variables is analogous to the convolution of their probability distributions.

quadratically with the number of gaussians.

According to [Bouy et al. \(2015\)](#), the number of Pleiades candidate members in the DANCe DR2 data set is 2010 (going up to 2109 in the combined Tycho-DANCe DR2). It means that the number of field objects dominates (99.9%) the DR2 data set. Even in our restricted 10^5 data set, RDR2 (see Sect. [2.7.2](#)), the field still dominates with a 0.98 fraction. Thus, it can be assumed that any new classification of candidate members will have a negligible impact on this figure. Therefore, it seems reasonable to assume that the GMM describing the field population can be frozen (fixed) during the process of cluster parameters inference. Let me elaborate on this assumption.

The objects that [Bouy et al. \(2015\)](#) classified as field are those whose cluster membership probabilities are lower than 0.75. This probability threshold corresponds to the one found by [Sarro et al. \(2014\)](#) after analysing the performance of their methodology when applied to synthetic data sets. In [Sarro et al. \(2014\)](#) the authors report that, at probability threshold $p = 0.75$, the contamination and true positive rates are $\approx 8\%$ and $\approx 96\%$ respectively. Assuming that these values are correct, the number of hypothetically misclassified objects is 12% of the cluster members, thus ≈ 241 objects. This value is negligible compared to the size of the data set (10^5 objects in the RDR2). It represents the negligible fraction of $\lesssim 0.0024$.

It can be further assumed that these hypothetically misclassified objects are spread in the observables space. Indeed, these misclassified can be thought to have membership probabilities near the classification threshold. Thus, they lay in the entanglement region which corresponds, in proper motions, to a halo around the cluster centre, and in photometry, to a region around the cluster sequence.

If the misclassified objects are a few and spread over the observable space, then their contribution to the parameters of the GMM describing the field population can be neglected. Thus, the parameters of the GMM remain fixed and out of the inference process.

The previous assumption is of paramount importance due to practical reasons. If I were to simultaneously infer the parameters of both cluster and field models, the required computing time will be excessive. The number of parameters in the field GMM goes up to ~ 300 , more than twice that of the cluster model (85 parameters). Instead, I first fit the parameters of the field GMM using the field objects in the RDR2 (the $\sim 98,000$ objects with membership probabilities below 0.75) and then I keep this parameters fixed in the inference process.

The number of gaussians for the proper motions and photometric GMM is found using the Bayesian Information Criterion (BIC, [Schwarz 1978](#)). The BIC is a model selection criterium that aims at avoiding over fitting. It represents a compromise between the likelihood, \mathcal{L} , of the n data points, and the number of parameters, k . This is,

$$BIC = \ln n \cdot k - 2 \ln \mathcal{L}. \quad (3.22)$$

To estimate the parameters of the GMM we use the Expectation Maximisation (EM) algorithm. However, the missing values in the photometry prevent the use of the standard from of the algorithm (see for example Chapter 9 of [Bishop 2006](#)). Instead, I estimated these parameters with the modified version of the EM algorithm for GMM found by [McMichael \(1996\)](#). On this version, objects with missing values also contribute to estimate the maximum-likelihood (ML) parameters. I applied this algorithm to GMM whose number of components ranged from 1 to 20. The optimal number of gaussians suggested by the BIC is 14.

In the case of proper motions, since they do not contain missing values, I computed the GMM parameters with the standard EM algorithm. However, the BIC favours models with large variances, large number of components, and small fractions. Since the proper motions distribution of field objects resembles a bivariate normal, the GMM suggested by BIC was contrary to our expectations. Therefore, we decided to add a uniform distribution to the mixture of Gaussians. This uniform distribution was intended to remove the large variance and small fraction components of the GMM. I changed the EM algorithm to properly account for this modification. The BIC applied to this new mixture of distributions renders more reasonable results. This modification improves the likelihood and reduces the number of parameters. The number of gaussians suggested by the BIC for this new mixture is 7, plus the uniform distribution.

Finally, the field likelihood $p_f(\mathbf{d}|\boldsymbol{\theta}_f, \mathbf{u})$ of an object with measurements \mathbf{d} , given the field parameters, $\boldsymbol{\theta}_f$, and the standard uncertainties \mathbf{u} of datum \mathbf{d} , is

$$p_f(\mathbf{d}|\boldsymbol{\theta}_f, \mathbf{u}) = [c \cdot \pi_{f,pm,0} + \\ \sum_{i=1}^7 \pi_{f,pm,i} \cdot \mathcal{N}(\mathbf{d}_{pm}|\boldsymbol{\mu}_{f,pm,i}, \boldsymbol{\Sigma}_{f,pm,i} + \mathbf{u}_{pm})] \\ \cdot \left[\sum_{i=1}^{14} \pi_{f,ph,i} \cdot \mathcal{N}(\mathbf{d}_{ph}|\boldsymbol{\mu}_{f,ph,i}, \boldsymbol{\Sigma}_{f,ph,i} + \mathbf{u}_{ph}) \right]. \quad (3.23)$$

In this equation, $\boldsymbol{\pi}_f, \boldsymbol{\mu}_f, \boldsymbol{\Sigma}_f$ represent, respectively, the fractions, means and covariance matrices of the GMM. The first and second brackets represent the proper motion and photometric models, respectively. The first term of the proper motion model is the uniform distribution. In it, c is a constant determined by the inverse of the product of the proper motions ranges (see Table 2.1), and $\pi_{f,pm,0}$ is the fraction of this uniform distribution. The second term in the same bracket is the mixture of gaussians with means $\boldsymbol{\mu}_{f,pm}$ and covariance matrices $\boldsymbol{\Sigma}_{f,pm} + \mathbf{u}_{pm}$.

3.3.4 The cluster population

Similarly to what I assume for the field population model, I also assume that the cluster model or likelihood can be factorised in the product of the proper motions distribution

times the photometric distribution. Thus, I assume these two models are independent.

It is known that unresolved systems of stars (groups of stars that, given the spatial resolution of the telescope, are seen as an individual object) have an increased brightness proportional to the multiplicity of the system. In particular, if an unresolved system is made of two equally luminous objects, then its magnitude is 0.752 times brighter than that of an individual object. This happens on equal mass binaries.

Since the pioneer work of [Trumpler \(1921\)](#), we know that some of the Pleiades members are double systems. Recently, [Sarro et al. \(2014\)](#) show evidence that this double systems lay in an equal-mass binaries (EMB) sequence. Those authors model this displaced sequence assuming that the number of objects on it is 20% of the total number of members. In the present work, we also model objects in this displaced EMB sequence. However, we do not assume that its proportion is fixed, we infer it from the data.

Unresolved multiple systems, and particularly binary systems have an impact on the cluster proper motions distribution. From an astrophysical point of view, in stellar clusters, the massive objects fall towards the centre of the gravitational potential in a rate higher than that of the less massive ones. This behaviour arises from stellar encounters in which the energy exchange results in the less massive objects gaining speed and the massive ones losing it. From an astronomical point of view, an unresolved system shifts the photo centre of its images when compared to that of a single object. Given the previous considerations, we decide to model the EMB as an independent population in the proper motions. Furthermore, we pair this proper motions model to its photometric one. This more comprehensive statistical model allow us to directly compare the kinematic and photometric properties of the EBM population with those of the rest of the cluster.

For the sake of simplicity, in the following whenever I refer to the photometric or proper motions model of the EMB I call it the EMB sequence model (with the subindex Bs). I refer to the model of the rest of the stars as the cluster sequence model or simply single stars. Although this is an abuse of the terminology, because there are binaries and multiple systems with different mass ratios, it keeps the text more readable.

3.3.4.1 Photometric model of equal-mass binaries and single stars

The cluster photometric sequences, both for single and EMB, are modelled with cubic splines, one for each of the $YJHK_s$ vs CI CMDs. I choose the spline series due to their fitting properties. I tried several polynomial bases (Laguerre, Hermite, Chebyshev) but in spite of the order, they were not able to fit the sequence, particularly in the region around $CI \approx 3$, where the slope is higher.

Despite their superior flexibility, spline series require more parameters than common polynomials. In addition to the coefficients of the series, they require the specification of points called knots. These knots represent the starting and ending points of the spline sections.

Any spline function can be represented in terms of basis-splines or B-splines. By definition, a B-spline of order n is a piece wise polynomial function of order n in the interval $t_0 \leq x \leq t_n$. The boundary and internal points, $\mathbf{t} = \{t_0, t_1, \dots, t_n\}$ represent the knots. For a given set of knots, there is one and only one B-spline representation of the spline, thus the name basis spline. In particular, any cubic spline can be represented as,

$$S_3(CI, \boldsymbol{\beta}, \mathbf{t}) = \sum_i \beta_i \cdot B_{i,3}(CI, \mathbf{t}). \quad (3.24)$$

Where $B_{i,n}$ are the B-splines given by the Cox-de Boor recursive formula, and $\boldsymbol{\beta}$ are the coefficients of the series. For more details on splines and the Cox-de Boor formula see [De Boor \(1978\)](#).

Despite their fitting properties, B-splines have an issue when inferring simultaneously their coefficients and knots: there is multi modality in the parametric space ([Lindstrom 1999](#)). It means that at least more than one combination of parameters produces the same solution. To avoid this multi modality, I decided to keep the knots fixed throughout the inference. This decision, although reduces the flexibility of the splines, allows a still better fit than that of the tested polynomials. To obtain the ML estimate of the knots I use the algorithm of [Spiriti et al. \(2013\)](#). This algorithm, implemented in the *freetknotsplines* R package, allows to simultaneously obtain the knots and the best truncation value for the spline series. It uses the BIC to select among competing models. In order to obtain both the truncation of the series and the value of the knots, I use the candidate members of [Bouy et al. \(2015\)](#). The BIC indicates that seven coefficients is the best number of components for the B-spline series, with the knots at $\mathbf{t} = \{0.8, 3.22, 3.22, 5.17, 8.0\}$. I tested different number of knots, ranging from two to nine, with five the best configuration given by the BIC. Since the knot at 3.22 is repeated it means that the resulting spline has lost a degree of continuity in this knot, its derivative is not continuous here.

As I mentioned in the introduction to this Section, we assume that the observed photometric quantities are drawn from a probability distribution resulting from the convolution of the observed uncertainties, with the *true* quantities. Here comes an expiation. I recognise that the model is far from perfect and that there are several astrophysical phenomena that it does not addresses. Either because I do not have the knowledge or because they are too complicated to model. These phenomena include, but are not limited to, age, metallicity and distance dispersions, unresolved systems (other than EMB), variability, and transits. Instead of independently modelling each one of them, I assume that all, even those that I ignore, contribute to the model in the shape of an intrinsic photometric dispersion. Given the large and heterogeneous origins of these phenomena, we can safely assume that its total contribution has the shape of a multivariate normal distribution. The parameters of this distribution can be inferred from the data as well.

We assume that this multivariate gaussian describes both the dispersion of the cluster

and EMB sequences. Most importantly, we distinguish this dispersion from the photometric uncertainty of individual measurements. If we were to assume no *true* intrinsic dispersion, then any deviation from the *true* quantities should have to be explained *only* by the observational uncertainty.

In our photometric model of the cluster, photometric observations are drawn from a multivariate normal distribution with five dimensions corresponding to our photometric reference set: CI, Y, J, H, K_s . The B-splines model the mean of the *true* photometric quantities, both for the cluster sequence, $\mathbf{t}_{ph;Cs}$, and the EMB, $\mathbf{t}_{ph;Bs}$. This later one is displaced 0.75 into the bright side of the cluster sequence. The matrix, Σ_{clus} , represents the covariance matrix of the multivariate normal distribution that models the intrinsic dispersion. Covariance matrices are symmetric and positive semi-definite. Therefore, from the 25 entries in this $5 \times 5 \Sigma_{clus}$ matrix, only 15 are unique. These are the ones we infer from the data.

Thus, the mean *true* photometry is given by,

$$\begin{aligned}\mathbf{t}_{ph;Cs} &= \{CI, Y, J, H, K_s\}, \\ \mathbf{t}_{ph;Bs} &= \{CI, Y - 0.75, J - 0.75, H - 0.75, K_s - 0.75\},\end{aligned}$$

where

$$\begin{aligned}Y &= \mathcal{S}_Y(CI, \beta_Y), \\ J &= \mathcal{S}_J(CI, \beta_J), \\ H &= \mathcal{S}_H(CI, \beta_H), \\ K_s &= \mathcal{S}_{K_s}(CI, \beta_{K_s}),\end{aligned}$$

and β_{Y,J,H,K_s} denote the coefficients of all the splines (a 4×7 matrix), which I call $\boldsymbol{\beta}$ for the sake of simplicity.

Since the photometry of the EMB is a linear transformation, T_{Bs} , of the mean *true* photometry of cluster sequence, no extra parameters are required. Therefore,

$$\mathbf{t}_{ph;Cs} = \mathcal{S}(CI, \boldsymbol{\beta}) \quad (3.25)$$

$$\mathbf{t}_{ph;Bs} = T_{Bs}(\mathcal{S}(CI, \boldsymbol{\beta})). \quad (3.26)$$

Thus, cluster and EMB likelihoods of an object with photometric measurements \mathbf{d}_{ph} ,

and standard uncertainties \mathbf{u}_{ph} , are:

$$\begin{aligned} p_{Cs}(\mathbf{d}_{ph}|CI, \boldsymbol{\beta}, \Sigma_{clus}, \mathbf{u}_{ph}) &= \mathcal{N}(\mathbf{d}_{ph}|\mathbf{t}_{ph;Cs}, \mathbf{u}_{ph} + \Sigma_{clus}), \\ p_{Bs}(\mathbf{d}_{ph}|CI, \boldsymbol{\beta}, \Sigma_{clus}, \mathbf{u}_{ph}) &= \mathcal{N}(\mathbf{d}_{ph}|\mathbf{t}_{ph;Bs}, \mathbf{u}_{ph} + \Sigma_{clus}), \end{aligned} \quad (3.27)$$

where $\mathbf{t}_{ph;Cs}$ and $\mathbf{t}_{ph;Bs}$ are given by Equations 3.25 and 3.26, respectively.

Since the splines are parametrised by the true CI of each object, we have more parameters than objects in our data set ¹¹. This *true CI* is unknown even if its observed value is not missing. We solve this problem (it is a computational problem!) by marginalising these nuisance parameters (See Eq. 3.29 and 3.30).

To marginalise these CIs we need a prior, which we provide in a hierarchical way (thus the name Bayesian Hierarchical model). This marginalisation leaves behind a precise estimate of the parameters of the prior distribution. Paradoxically, all objects, even those without a measurement of the CI contribute to this estimate. Here lays the force of the BHM.

We model the prior of the *true CI* as a truncated ($0.8 \leq CI \leq 8$) univariate GMM with five components, whose parameters are also inferred from the data. We chose five components as suggested by BIC. I applied BIC to the results found using the standard EM algorithm for univariate GMM applied to the candidate members of Bouy et al. (2015). I tested with larger number of components in the mixture, but the posterior distribution did not changed, thus indicating that the BIC value was a proper assumption.

Thus, the GMM modelling the prior of the *true CI* is

$$p_{CI}(CI|\boldsymbol{\pi}_{CI}, \boldsymbol{\mu}_{CI}, \boldsymbol{\sigma}_{CI}) = \sum_{i=1}^5 \pi_{CI,i} \cdot \mathcal{N}_t(CI|\mu_{CI,i}, \sigma_{CI,i}). \quad (3.28)$$

In this last Equation, the symbol \mathcal{N}_t stands for the truncated ($0.8 < CI < 8$) univariate normal distribution.

Then, the marginalisation of CI runs as follows:

$$\begin{aligned} p_{Cs}(\mathbf{d}_{ph}|\boldsymbol{\theta}_c, \mathbf{u}_{ph}) &= \int p_{Cs}(\mathbf{d}_{ph}, CI|\boldsymbol{\theta}_c, \mathbf{u}_{ph}) \cdot dCI \\ &= \int p_{Cs}(\mathbf{d}_{ph}|CI, \boldsymbol{\theta}_c, \mathbf{u}_{ph}) \cdot p_{Cs}(CI|\boldsymbol{\theta}_c, \mathbf{u}_{ph}) \cdot dCI \end{aligned} \quad (3.29)$$

$$\begin{aligned} p_{Bs}(\mathbf{d}_{ph}|\boldsymbol{\theta}_c, \mathbf{u}_{ph}) &= \int p_{Bs}(\mathbf{d}_{ph}, CI|\boldsymbol{\theta}_c, \mathbf{u}_{ph}) \cdot dCI \\ &= \int p_{Bs}(\mathbf{d}_{ph}|CI, \boldsymbol{\theta}_c, \mathbf{u}_{ph}) \cdot p_{Bs}(CI|\boldsymbol{\theta}_c, \mathbf{u}_{ph}) \cdot dCI. \end{aligned} \quad (3.30)$$

In these Equations, $\boldsymbol{\theta}_c$ stands for all cluster parameters related to photometry, and the

¹¹Although this sounds crazy, the rules of probability calculus do not discard this possibility.

first and second terms of the integrals in the last equalities correspond to Equations 3.27 and 3.28, respectively. The distribution of CI depends only on $\boldsymbol{\pi}_{CI}$, $\boldsymbol{\mu}_{CI}$, $\boldsymbol{\sigma}_{CI}$, thus, the cluster and equal-mass binaries likelihoods of datum \mathbf{d}_{ph} are

$$\begin{aligned} p_{Cs}(\mathbf{d}_{ph} | \boldsymbol{\pi}_{CI}, \boldsymbol{\mu}_{CI}, \boldsymbol{\sigma}_{CI}, \boldsymbol{\beta}, \Sigma_{clus}, \mathbf{u}_{ph}) \\ = \int \mathcal{N}(\mathbf{d}_{ph} | \mathcal{S}(CI, \boldsymbol{\beta}), \mathbf{u}_{ph} + \Sigma_{clus}) \\ \cdot \sum_{i=1}^5 \pi_{CI,i} \cdot \mathcal{N}_t(CI | \mu_{CI,i}, \sigma_{CI,i}) \cdot dCI \\ p_{Bs}(\mathbf{d}_{ph} | \boldsymbol{\pi}_{CI}, \boldsymbol{\mu}_{CI}, \boldsymbol{\sigma}_{CI}, \boldsymbol{\beta}, \Sigma_{clus}, \mathbf{u}_{ph}) \\ = \int \mathcal{N}(\mathbf{d}_{ph} | T_{Bs}(\mathcal{S}(CI, \boldsymbol{\beta})), \mathbf{u}_{ph} + \Sigma_{clus}) \\ \cdot \sum_{i=1}^5 \pi_{CI,i} \cdot \mathcal{N}_t(CI | \mu_{CI,i}, \sigma_{CI,i}) \cdot dCI. \end{aligned} \quad (3.31)$$

The observed CI and magnitudes help us to reduce the computing time of the marginalisation integral. We use them to discard regions of the integral in which the argument is almost zero (i.e. far from the measured values). Although we allow the nuisance parameters CIs to have all their possible values, the data, by means of the likelihood, give us information about the distribution of these individual nuisance parameters. To use this information, we proceed as follows. First, we compare the observed photometry to the true one (i.e. the cluster sequence given by the splines). For it we use a grid of 300 points uniformly distributed in the domain of CI ($0.8 < CI < 8$). Then, we find the point, p , of the grid that is the closest to the vector of the observed photometry. Distance is computed under the Mahalanobis metric. This metric takes into account the observational uncertainty, \mathbf{u}_{ph} , and the intrinsic dispersion of the cluster sequence, Σ_{clus} . Finally, the limits of the marginalisation integral are defined as those given by a ball of 3.5 Mahalanobis distances around point p . Contributions outside this ball are negligible to the integral ($< 4 \times 10^{-4}$).

3.3.4.2 Proper motion model of equal-mass binaries and single stars

As mentioned before, we assume that the cluster population has two subpopulations: single and EMB stars. We model the proper motions of these two subpopulations with independent GMM. If the cluster is virialised (see Chapter 2), we can assume that its velocity distribution is almost Maxwellian (Maxwell-Boltzman distribution). Therefore a GMM is a reasonable assumption. Furthermore, in the absence of external forces, a virialised system is expected to have spherical symmetry both in its spatial and velocity distributions. Thus we can safely assume that the gaussians within each GMM are concentric, thus they share the same mean. However we allow independent means for both

single and EMB subpopulations. The assumption of spherical symmetry may be a weak one in the presence of the galactic potential. It can perturb the cluster and deviate its spatial and velocity distribution from spherical symmetry. Furthermore, the ellipticity of the spatial distribution, which has been reported to be no negligible ($\epsilon = 0.17$, according to [Raboud & Mermilliod 1998](#)), can due to projection effects further deviate the observed velocity distribution profile from spherical symmetry. Nevertheless, since we model the covariance matrices of the GMM of both single and EMB, as full covariance matrices, any departure from the spherical symmetry in the velocity distribution can still be perfectly modelled by the non diagonal entries of these matrices.

We infer the parameters of these GMMs as part of our Bayesian hierarchical model. However we a priori set the number of gaussians in each GMM. Not doing so will demand a technique in which the model parameters can be augmented. Although such techniques already exist, they are still under computational development (see [Fan & Sisson 2011](#), for a review of reversible jump MCMC).

Using the EM for GMM and the proper motions of the candidate members of [Bouy et al. \(2015\)](#), I obtained the ML estimates for the GMM likelihoods. I did this for configurations of GMM ranging from one to five components. The BIC (Eq. 3.22) suggested four and two gaussians for cluster and EMB GMM, respectively. Since covariance matrices are always symmetric, only three parameters are needed to fully specify the covariance matrices of these bivariate normal distributions.

Thus, the cluster (subindex Cs) and EMB (subindex Bs) likelihoods of an object with proper motions measurements \mathbf{d}_{pm} , and uncertainties \mathbf{u}_{pm} , are

$$\begin{aligned} p_{Cs}(\mathbf{d}_{pm}|\boldsymbol{\pi}_{Cs}, \boldsymbol{\mu}_{Cs}, \boldsymbol{\Sigma}_{Cs}, \mathbf{u}_{pm}) &= \sum_{i=1}^4 \pi_{Cs,i} \cdot \mathcal{N}(\mathbf{d}_{pm}|\boldsymbol{\mu}_{Cs}, \Sigma_{Cs,i} + \mathbf{u}_{pm}) \\ p_{Bs}(\mathbf{d}_{pm}|\boldsymbol{\pi}_{Bs}, \boldsymbol{\mu}_{Bs}, \boldsymbol{\Sigma}_{Bs}, \mathbf{u}_{pm}) &= \sum_{i=1}^2 \pi_{Bs,i} \cdot \mathcal{N}(\mathbf{d}_{pm}|\boldsymbol{\mu}_{Bs}, \Sigma_{Bs,i} + \mathbf{u}_{pm}). \end{aligned} \quad (3.32)$$

Finally, combining the proper motions and photometric models, the total cluster likelihood of an object with measurement \mathbf{d} , and uncertainties \mathbf{u} , is

$$\begin{aligned} p_c(\mathbf{d}|\boldsymbol{\theta}_c, \mathbf{u}) &= \pi_{CB} \cdot p_{Cs}(\mathbf{d}_{pm}|\boldsymbol{\pi}_{Cs}, \boldsymbol{\mu}_{Cs}, \boldsymbol{\Sigma}_{Cs}, \mathbf{u}_{pm}) \\ &\quad \cdot p_{Cs}(\mathbf{d}_{ph}|\boldsymbol{\pi}_{CI}, \boldsymbol{\mu}_{CI}, \boldsymbol{\sigma}_{CI}, \boldsymbol{\beta}, \Sigma_{clus}, \mathbf{u}_{ph}) \\ &\quad + (1 - \pi_{CB}) \cdot p_{Bs}(\mathbf{d}_{pm}|\boldsymbol{\pi}_{Bs}, \boldsymbol{\mu}_{Bs}, \boldsymbol{\Sigma}_{Bs}, \mathbf{u}_{pm}) \\ &\quad \cdot p_{Bs}(\mathbf{d}_{ph}|\boldsymbol{\pi}_{CI}, \boldsymbol{\mu}_{CI}, \boldsymbol{\sigma}_{CI}, \boldsymbol{\beta}, \Sigma_{clus}, \mathbf{u}_{ph}), \end{aligned} \quad (3.33)$$

where π_{CB} is the parameter representing the amplitude of single cluster sequence stars in the single-EMB mixture model. The photometric and proper motions likelihoods are

given by Equations 3.31 and 3.32, respectively.

Finally, before ending this Section, I present a summary of the symbols of the parameters in the BHM (Table 3.1), together with its graphical representation in the form of a PGM (Fig. 3.3).

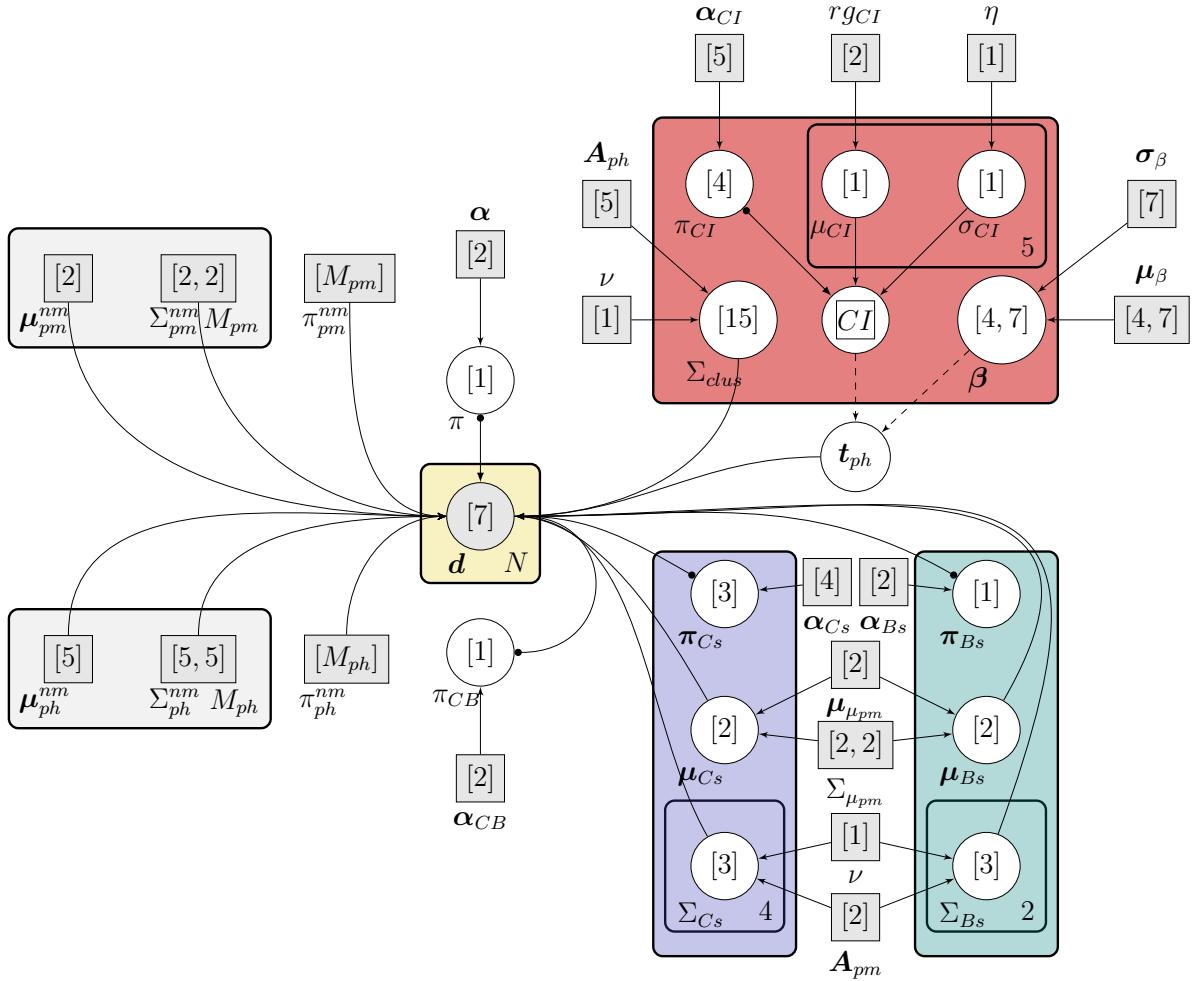


Figure 3.3: Probabilistic graphical model representing the BHM. The left grey plates show the field model. The middle yellow plate shows the node where the likelihood is computed for each datum, \mathbf{d} . The right plates describe the relations among parameters in the cluster model. The photometric cluster model (red) is on top, while the proper motions cluster (blue) and equal-mass binaries (green) are at the bottom left and right, respectively. See Section 3.2.3 for more details. Reproduced from Figure 19 of Olivares et al. (2017), *The Seven sisters DANCE III: Hierarchical Bayesian model*, A&A, submitted.

3.4 Priors

The Bayesian formalism is characterised by the use of priors. These priors, as mentioned earlier, represent the objective (measurable and reproducible) way to establish beliefs about the distribution of values that certain parameter may have.

In the following, I describe the information used to establish both the family of the prior distribution as well as its hyper-parameters (the parameters at the top hierarchy of the BHM). As mentioned before, these families are chosen to fall, whenever possible, in the category of weakly informative priors.

The priors in the BHM can be grouped into three main categories. The first one correspond to priors for parameters representing fractions in mixture models. The second and third categories correspond to parameters in the proper motions and photometric models.

In the BHM, there are different types of mixtures: the GMMs of the proper motions, the cluster-field mixture, and the singles-EMB mixture. At each mixture, the fractions or amplitudes quantify the contribution of each element in the mixture to the probability distribution. Fractions must add to one and be bounded by the $[0, 1]$ interval. We choose the Dirichlet distribution to be the family of all fraction parameters. This decision roots in the fact that this distribution is the multivariate generalisation of the beta distribution. This last one is commonly used to model the probability of success of an event.

The Dirichlet distribution has support in $[0, 1]$ for each entry, and is parametrised by $\boldsymbol{\alpha}$. This distribution gives the probability of n rival events given that each rival event has been observed $\alpha_i - 1$ times ($i = \{1, 2, \dots, n\}$). The mean and variance of the Dirichlet distribution are given by,

$$E[x_k] = \frac{\alpha_k}{\sum_k \alpha_k}, \quad (3.34)$$

$$Var[x_k] = \frac{-\alpha_k \cdot (\alpha_k - \sum_k \alpha_k)}{(\sum_k \alpha_k)^2 \cdot (1 + \sum_k \alpha_k)}. \quad (3.35)$$

For the field-cluster mixture we set the hyper-parameters to $\boldsymbol{\alpha} = \{9.8, 0.2\}$. We expect a mean 98% of field objects and a 2% of cluster objects with little variance. These figures correspond to the existent prior knowledge, this is the fraction of field and cluster candidate members found by [Bouy et al. \(2015\)](#). For the single-EMB mixture we use an hyper-parameter value, $\boldsymbol{\alpha}_{CB} = \{8, 2\}$. We expect a mean 20% of EMB, as suggested by [Bouy et al. \(2015\)](#). For fractions in the proper motions GMM, hyper-parameter are $\boldsymbol{\alpha}_{Cs} = \{1, 1, 5, 5\}$ and $\boldsymbol{\alpha}_{Bs} = \{1.2, 8.8\}$. These values induce fraction distributions whose means are similar to the fractions recovered after fitting a GMM to the candidate members of [Bouy et al. \(2015\)](#). For the fraction in the GMM of the *CI* distribution, the hyper-parameter were set all to 1, ($\boldsymbol{\alpha}_{CI} = \{1, 1, 1, 1, 1\}$), which results in equal means and large variances for all of them.

In the previous cases, with exception of the cluster-field mixture, the hyper parameters were chosen to have larger variances. Fig. 3.4 shows the distribution associated to these hyper-parameters.

The narrow variance in the cluster-field mixture expresses our prior belief about the

number (fraction) of candidate members within our large data set.

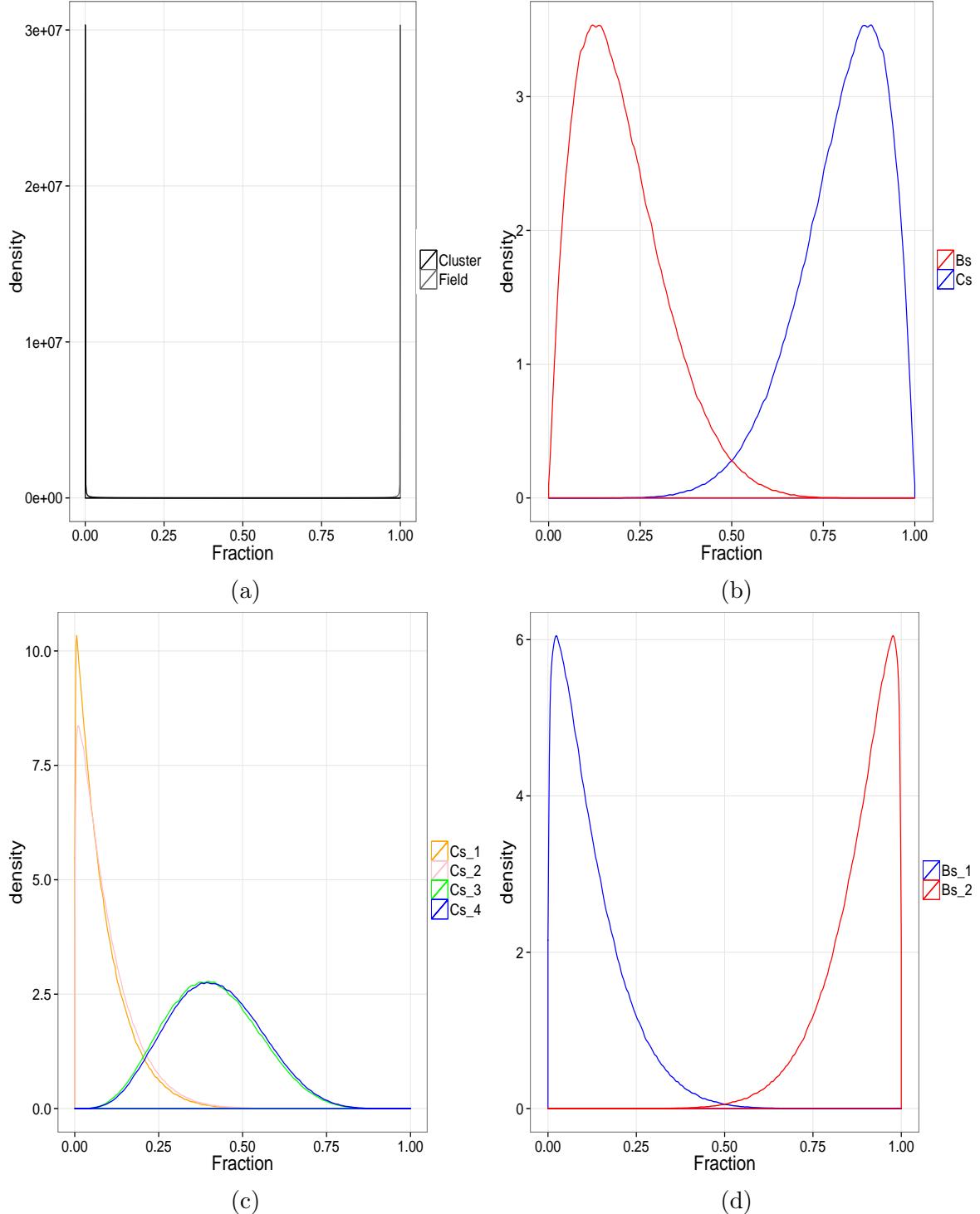


Figure 3.4: Prior distributions of fraction parameters. Distributions of: (a) The field fraction π , (b) The EMB fraction, $1 - \pi_{CB}$, (c) the proper motions cluster fractions, π_{Cs} , and (d) the proper motions equal-mass binaries fractions π_{Bs} .

For the priors of mean parameters in the proper motions GMM, both of single and EMB, we choose the bivariate normal distribution. We set the hyper-parameters of this bivariate normal to those found after fitting a bivariate normal to the candidate members of Bouy et al. (2015). These values are $\mu_{\mu_{pm}} = \{16.30, -39.62\}$ and $\Sigma_{\mu_{pm}} =$

$$\{\{36.84, 1.18\}, \{1.18, 40.71\}\}.$$

As prior for the covariance matrices of both single and EMB proper motions GMM we use the Half- $t(\nu, \mathbf{A})$ distribution. It is parametrised by an scalar ν and a vector \mathbf{A} . As shown by [Huang & Wand \(2013\)](#) this family distribution leads to more accurate estimation of covariance matrices than the traditional Inverse-Wishart distribution. Particularly, the marginal correlation correlation parameters, ρ have the following distribution,

$$p(\rho) \propto (1 - \rho^2)^{\frac{\nu}{2}-1}. \quad (3.36)$$

The standard deviation term σ_k , associated to entry k , is distributed according to Half- $t(\nu, A_k)$. We set the hyper-parameter to $\nu = 3$ and $\mathbf{A}_{pm} = \{10^5, 10^5\}$. According to [Huang & Wand \(2013\)](#), an arbitrarily large values of \mathbf{A} lead to arbitrarily weakly informative priors on the corresponding standard deviation terms.

Concerning the photometric priors, they can be grouped in three categories: (i) priors for the the *true CI*, (ii) priors for the splines coefficients, and (iii) priors for the cluster sequence intrinsic dispersion.

For the means in the univariate GMM modelling the *true CI* I choose a uniform distribution in the range $(0.8 \leq CI \leq 8)$. For the standard deviations I choose the Half-Cauchy($0, \eta$) distributions as suggested by [Gelman \(2006\)](#). The value of η is set to an arbitrarily large value. In this case I chose $\eta = 100$.

For the coefficients in the spline series we set the priors as univariate normal distributions. To find the mean and variance of these distributions we proceed as follows. First, we remove the EMB from the list of candidate members of [Bouy et al. \(2015\)](#). To do this, I performed an iterative fit of the cluster sequence, at each iteration I removed those objects whose photometry was as bright as that of the EMB. In the region of $CI > 7$ there are no candidate members of [Bouy et al. \(2015\)](#) and of any other source. Thus, to provide a prior we complement our list of candidate members with the brown-dwarfs from the [Faherty et al. \(2012\)](#) sample. We chose only those objects observed in the same photometric bands of our data set. Finally, we fit the splines, and use the coefficients of this fit as the means, μ_β of the univariate normal distributions. The standard deviation terms were set to $\sigma_\beta = \{1, 1, 1, 1, 1, 0.5, 0.1\}$. These values provide a reasonable compromise between cluster sequences compatible with the previously known candidates, and those far away or with exotic shapes. We show a sample of this priors in Fig. 3.5. This Fig. also shows the brown-dwarfs from [Faherty et al. \(2012\)](#) and the sequence (dashed line) we use to provide the means of the univariate normal distributions.

To set the prior for the parameters of the cluster intrinsic dispersion, Σ_{clus} , I choose again the Half- $t(\nu, \mathbf{A})$ distribution. However, this time I use $\mathbf{A}_{ph} = \{10, 10, 10, 10, 10\}$. These values are large when compared to the standard deviation terms of the observed uncertainties. Therefore, this values provide a weakly informative prior on the marginal standard deviation terms of the Σ_{clus} covariance matrix.

Table 3.2 shows a summary all the hyper-parameter and their values.

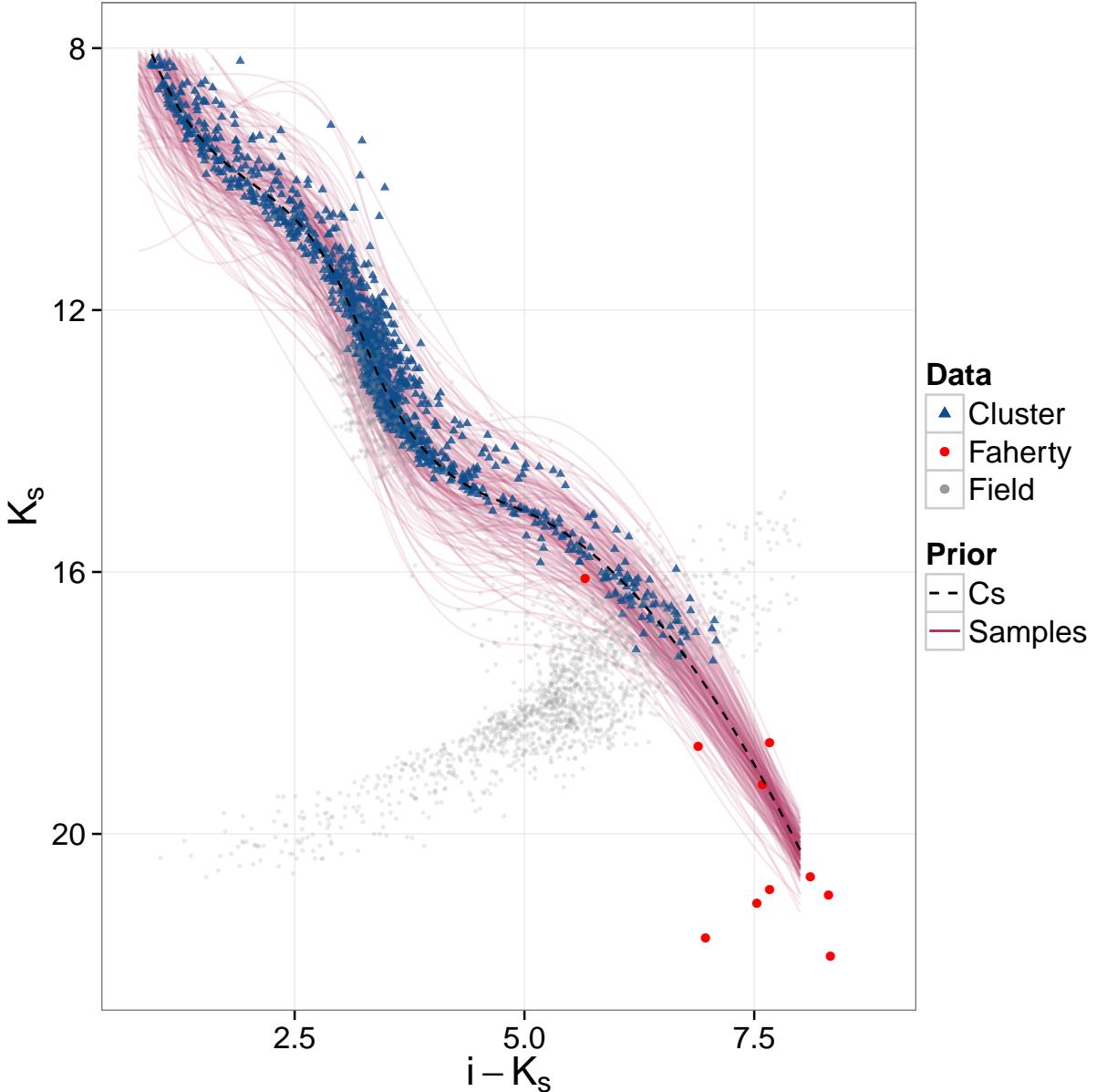


Figure 3.5: CMD K_s vs. $i - K_s$ showing a sample (100 elements) of the prior for the coefficients in the splines series. Also shown are the brown-dwarfs from [Faherty et al. \(2012\)](#) sample (red dots), the cluster sequence (dashed line) used as mean for the priors, the candidate members of [Bouy et al. \(2015\)](#) (blue triangles), and a sample of the field (grey dots).

3.5 Sampling the posterior distribution

Theoretically, there are at least three possible approaches to obtain the posterior distributions of the parameters in our model. One of these options is the analytical approach. Analytically computing the likelihood of our large data set is not feasible, thus I

discard this approach. The second option is the use of a grid in the parametric space. The likelihood and the prior must be evaluated at each point in this grid and then multiplied. This approach is reasonable when the parametric space is of moderate dimension (≤ 5). It requires the evaluation of the posterior distribution q^p times, with q the number of grid points in one dimension, and p the dimension of the parametric space. The number of parameters in our model is 85, which immediately rules out this possibility. The third and so far only feasible approach is the use of Markov Chain Monte Carlo (MCMC) methods. Although these methods provide a solution in a reasonable time, nevertheless, the bottle neck of computing time is due to the evaluation of the likelihood, which grows linearly with the size of the data set.

This Section is structured as follows. First, I introduce an heuristic technique to perform a fast search of the maximum a posteriori of our target distribution. Then, I describe the MCMC techniques available in the literature. Particularly I focus the one technique we chose, and in the reasons for which it was chosen. The section ends with details about the assessment of the MCMC convergence.

3.5.1 PSO

The likelihood of the data is the product of the individual likelihoods of each datum (Eq. 3.10). Therefore, the number of operations needed to evaluate the likelihood grows proportionally to the size of the data set. As I will explain in Section 3.5.2, the burn-in phase of MCMC techniques allows them to reach the target distribution. However, once the MCMC reaches this target distribution, the burn-in computations are discarded. Since the evaluation of the likelihood, and therefore of the posterior, is computationally expensive, I decided to reduce as much as possible the burn-in phase. To do so, I provide MCMC with a set of initial solutions which are close to the Maximum-A-Posteriori (MAP) of the target (posterior) distribution. These near-MAP solutions must not be too crowded on the MAP, otherwise MCMC will spend time in expanding them to reach the entire distribution. Here, there is a trade off between the crowdedness of the solutions and its proximity to the MAP. Since we aim at obtaining a representative sample of the posterior distribution and not just an estimate of it, then the initial set of solution for the MCMC must be carefully chosen to minimise the computing time. This section provides the details of this procedure.

In Section 3.5.2, I will also show that the MCMC flavour more suitable to our objective belongs to the family of *ensemble* MCMC. This flavour works with particles in the parametric space. To make the transition between the initial MAP solutions and the MCMC particles as efficient as possible, I chose a MAP finder which works also with particles: the Particle Swarm Optimiser (PSO, [Kennedy & Eberhart 1995](#)). I provides a heuristic cheap and fast approach to the MAP solution. The PSO works with an ensemble of particles which move through the parametric space. These particles use the collective

and individual past and present information to update their position. This information is specified by the score function, which in our case is the posterior distribution. The particles update their position iteratively by means of their velocity. This velocity has a random but restricted magnitude, however, its direction is determined by the particle position, and the individual and collective positions with maximum score. [Kennedy & Eberhart \(1995\)](#) has a detailed description of the original algorithm, while a more efficient version is given by ([Clerc & Kennedy 2002](#)).

Although the PSO is a simple and rather efficient solution to the MAP approximation, it is far from perfect. Due to its heuristic origin, there is no theory behind its formulation. Furthermore, it does not always guarantee the finding of the global maximum (for a convergence guaranteed version see [Patel et al. 2013](#)). Although, this issue does not affect our results (as we will see MCMC does guarantee the finding of the target distribution), it impacts the computing time. If the global maximum is not found in the PSO stage, then the MCMC will take longer to arrive to the target distribution.

On the other hand, the PSO stops its computations once the mean of the particles scores lays within a user defined tolerance. If this tolerance is too high, the PSO may stop far from the MAP. If it is too small, it may converge to the MAP but deliver solutions highly concentrated around it. This poses a problem to the following MCMC stage. For it to explore the full posterior distribution, the MCMC will need more iterations, thus more time, to expand the initially concentrated positions. The optimal value for the relative tolerance is 10^{-7} . I found this after trial and error.

To overcome the problem of crowdedness, I decided to use the charged PSO ([Blackwell & Bentley 2002](#)). Originally designed to optimise a time varying score function, the charged PSO maintains its exploratory capabilities due to an electrostatic force that repels particles when they get closer than a certain distance ([Blackwell & Bentley 2002](#)). Thanks to this electrostatic force the charged PSO avoids the over-crowding of particles around local best values.

The algorithm of [Blackwell & Bentley \(2002\)](#) computes distances in the entire parametric space. I find this approach unsuitable to our problem, thus I modified it. This modified version, together with the values chosen for its parameters is described in more details on Section [3.6.1](#).

3.5.2 MCMC

3.5.2.1 Generalities

Markov chain Monte Carlo (MCMC) is the generic name for a series of algorithms whose objective is the sampling of probability distributions. As their name indicates, the MCMC generates a chain (or a group of them) of Monte Carlo realisations that fulfil the Markov property. Monte Carlo realisations can be understood, broadly speaking as

continuous random realisations. Since it is an iterative algorithm, the chain is a process that refers to the joint of all random Monte Carlo steps. The Markov property indicates the probabilistic independence between steps in the chain that are separated more than one iteration. Thus, in a Markov chain, the probability of a future step depends only on the present step, and not in the past steps.

[Andrieu et al. \(2003\)](#) provides a brief and interesting summary of the history of the MCMC methods. In the following I use their work to describe the fundaments of MCMC. For more details, see the aforementioned authors and the book of [Brooks et al. \(2011\)](#).

A stochastic process is defined as a sequence $\{\theta_1, \dots, \theta_n\}$ of random elements. On it each element $\theta_i \subset \mathbb{R}^k$, with k the dimension of the *state space*.

A stochastic process, $\boldsymbol{\theta} = \{\theta_0, \theta_1, \dots, \theta_n, \theta_{n+1}\}$ is called a Markov chain if

$$p(\theta_{n+1} | \theta_0, \theta_1, \dots, \theta_n) = p(\theta_{n+1} | \theta_n).$$

A Markov chain has two important distributions, the initial distribution and the transition distribution. The initial distribution is the marginal distribution of θ_0 , it is $p(\theta_0)$. The transition distribution is the conditional probability $p(\theta_{n+1} | \theta_n)$. This last one is called stationary or homogeneous if it does not depend on n .

If this transition is irreducible and aperiodic, then there is an *invariant* or *equilibrium* distribution to which the chain converges to, in spite of the initial distribution. Here, aperiodic means that the chain does not make loops, whereas it is irreducible if the probability of exploring all other states is not zero.

If we want to have $p(\theta)$ as the invariant distribution, then it suffices that the transition distribution $p_t(\cdot | \cdot)$ satisfies the detailed balance condition,

$$p(\theta_n) \cdot p_t(\theta_{n-1} | \theta_n) = p(\theta_{n-1}) \cdot p_t(\theta_n | \theta_{n-1}). \quad (3.37)$$

Thus MCMC are Markov chains that satisfy the detailed balance condition, and had their invariant distribution as the target distribution. The large variety of MCMC algorithms arises from the efficiencies in which they arrive to the target distribution.

In the following I will review three of the MCMC categories: Metropolis-Hasting (MH), Hamiltonian Monte Carlo (HMC), and affine invariant samplers. The MH category comprises the classic MH algorithm but also contain particular cases like the Gibbs sampler ([Geman & Geman 1984](#)). I describe MH only for completeness and explanatory reasons. Later, I will focus on the particular cases of Hamiltonian Monte Carlo, for (HMC), affine invariant for ensemble samplers. Finally, I will briefly describe Nested Sampling, an algorithm that uses MCMC to numerically compute the Bayesian evidence and simultaneously samples the posterior distribution.

3.5.2.2 Metropolis-Hastings

By far, the most popular MCMC algorithm is Metropolis-Hastings ([Metropolis et al. 1953](#); [Hastings 1970](#)). Once the Markov chain has been initialised in the state space, given the current, θ , and the proposed, $\hat{\theta}$, positions, the chain moves from θ to $\hat{\theta}$ with acceptance probability:

$$\mathcal{A}(\hat{\theta}|\theta) = \min \left\{ 1, \frac{p(\hat{\theta}) \cdot q(\theta|\hat{\theta})}{p(\theta) \cdot q(\hat{\theta}|\theta)} \right\}. \quad (3.38)$$

Where q is the transition probability. Since the algorithm allows rejection, it is aperiodic, and to ensure irreducibility, the support of q must include that of p ([Andrieu et al. 2003](#)). The popularity of MH lies in its simplicity. Nevertheless it requires a careful tuning of the transition probability. Usually, this probability is given by a normal distribution. It works well for relatively low dimensions of the parametric space (≤ 5). However, once the dimension goes higher, the MH algorithm spend a great amount of time in tuning the parameters of this multivariate normal, particularly of its covariance matrix.

3.5.2.3 Hamiltonian Monte Carlo

The Hamiltonian Monte Carlo algorithms ([Duane et al. 1987](#); [Neal 1996](#)), as they name suggests¹², use Hamiltonian dynamics to express the target distribution as the potential distribution of a hamiltonian system of particles. In such systems the total energy is the sum of the potential and kinetic energies. The potential distribution depends only on position, whereas the kinetic one on momentum. HMC introduces a momentum to the particles in order to use their positions as a sample of the target distribution. To update the particles positions, HMC uses the Hamilton equations, which contain information about the gradient of the potential. Once HMC has tuned the momentum distribution, the proposed positions are more likely in terms of the target distribution. Therefore, using the information about the gradient of the target distribution, HMC is able to improve the acceptance ratio of the proposed steps. A detailed description of HMC can be found in Chapter 5 of [Brooks et al. \(2011\)](#). The package *Stan* ([Carpenter et al. 2017](#)) provides an efficient implementation of HMC.

3.5.2.4 Affine invariant

Affine invariant MCMC samplers use many particles, (ensemble), to sample the target distribution with a performance that is independent of its shape in the parametric space. Affine invariant MCMC do not need to tune the transition distribution, for this reason, these samplers are faster than standard MCMC ([Goodman & Weare 2010](#)). In the following I use the derivation of [Goodman & Weare \(2010\)](#).

¹²Originally called Hybrid Monte Carlo by ([Duane et al. 1987](#))

An ensemble $\boldsymbol{\theta}$, consists of a set of L particles $\theta_l \subset \mathbb{R}^k$ living in state space \mathbb{R}^{kL} . These particles are independently drawn from the target distribution π . Therefore,

$$\Pi(\boldsymbol{\theta}) = \pi(\theta_1) \cdot \pi(\theta_2) \dots \pi(\theta_L).$$

Thus, an ensemble MCMC is a Markov chain in the state space of ensembles, more properly, in the state space of the sequence $\boldsymbol{\theta}(1), \boldsymbol{\theta}(2), \dots, \boldsymbol{\theta}(t)$. An ensemble MCMC preserves the equilibrium distribution without the individual particles sequence, $\theta_1(1), \theta_1(2), \dots, \theta_1(t)$, being Markov or even independent. However, to update the particles positions, the detailed balance condition (Eq. 3.37) must be fulfilled. [Goodman & Weare \(2010\)](#) use partial resampling to ensure this. On it, the transition probability preserves the target (invariant) distribution if the single particle steps preserve the conditional distribution of the particle position given the complementary ensemble (the rest of the particles). Using the affine invariant *stretch move*, these authors are able to define a Markov chain, in the state space of ensembles, that satisfies the detailed balance condition.

The stretch move $\theta_k(t) \rightarrow \hat{\theta}$, defined as,

$$\hat{\theta} = \theta_j(t) + z \cdot (\theta_k(t) - \theta_j(t)),$$

where $\theta_j(t)$ is the current position of a particle in the complementary ensemble, and z is the stretching factor. It produces a symmetric transition, $p(\theta_k(t) \rightarrow \hat{\theta}) = p(\theta_k(t) \leftarrow \hat{\theta})$, if its density $g(z)$ satisfies the symmetry condition

$$g\left(\frac{1}{z}\right) = z \cdot g(z).$$

Finally, [Goodman & Weare \(2010\)](#) define their affine invariant MCMC using the following distribution for $g(z)$,

$$g(z) \propto \begin{cases} \frac{1}{\sqrt{z}} & \text{for } z \in [1/a, a] \\ 0 & \text{for } z \notin [1/a, a] \end{cases} \quad (3.39)$$

and the acceptance probability,

$$\mathcal{A}(\hat{\theta}|\theta) = \min \left\{ 1, z^{n-1} \cdot \frac{p(\hat{\theta})}{p(\theta)} \right\}. \quad (3.40)$$

The parameter a , which must be greater than one, improves the performance of the sampler ([Goodman & Weare 2010](#)).

One of the great advantages of ensemble samplers MCMC is its possibility of parallelisation. Since they work with particles, these particles can be distributed among cores in a computer cluster, therefore reducing the computing time when compared to non ensemble MCMC. [Foreman-Mackey et al. \(2013\)](#) implemented the affine invariant stretch move of

[Goodman & Weare \(2010\)](#) in the Python package *emcee*.

3.5.2.5 Nested sampling

Nested sampling ([Skilling 2004, 2006](#)) is an algorithm designed to numerically integrate the evidence (Eq. 3.11). As a by product it also delivers a sample of the posterior distribution. To compute the evidence integral, it uses particles positions sampled from the prior. Then, the integral is the sum of the likelihood, L_i , of each particle times a weight, w_i . This weight is a proxy for the volume of the prior covered by the updated position of the particle. This weight is given by $w_i = \exp(-(i-1)/N) - \exp(-i/N)$. Finally, the particles update their position only when their likelihood is the minimum of the group of particles. In this way,

$$z \leftarrow \sum_i w_i \cdot L_i \quad (3.41)$$

An improved version of the original algorithm was implemented in *MultiNest* ([Feroz et al. 2009](#)). This version allows the sampling and computing of evidence in the even more difficult multimodal posteriors.

3.5.3 Implementation and convergence: PSO and MCMC

To sample the posterior distribution in our problem, we chose *emcee* due to the following properties: i) the affine invariance allows a faster convergence over common and skewed distributions (see [Goodman & Weare 2010; Foreman-Mackey et al. 2013](#), for details), ii) it runs on parallel by distributing particles over nodes of a computer cluster, this reduces considerably the computing time, and iii) it requires the hand-tuning of only two constants: the number of particles, and the parameter a of the $g(z)$ distribution(Eq. 3.39). I choose a ratio of particles to parameters of two, it results in 170 particles. This is the minimum ratio recommended by [Foreman-Mackey et al. \(2013\)](#), which still allows a reasonable computing time. After trial and error, I fix the value of parameter a to $a = 1.3$. As mentioned by [Goodman & Weare \(2010\)](#) this parameter can be tuned to improve performance of the sampler. This value keeps the acceptance fraction in the range $0.2 - 0.5$, as recommended by [Foreman-Mackey et al. \(2013\)](#).

As a front-end of *emcee*, and to handle the input and output of data, I use a modified version of *CosmoHammer* ([Akeret et al. 2013](#)), (the next Section provides the details of these modifications).

As mentioned earlier, the PSO does not guarantee the finding of the global maximum of the score function. Therefore, I implement an iterative approach that minimises the risk of PSO to get stuck in local maxima. To do so, I iteratively run PSO and 50 iterations of *emcee* (with the same number of particles as the PSO) until the relative difference between

means of consecutive iterations is lower than 10^{-7} . The iterations of *emcee* spread the PSO solution without moving it away from the target distribution.

Neither scheme, PSO alone or PSO-*emcee*, guarantees to finding of the global maximum. The solution this scheme provides could indeed be biased. However, we use them to obtain a fast estimate of the global maximum, or at least, of points in its vicinity. If the initial solution provided by this scheme is indeed biased, the final *emcee* run erases, during the burning phase, any dependance on the biased initial solutions. After the convergence of the PSO-*emcee* scheme, I run *emcee* until it converges.

Convergence to the target distribution occurs when each parameter enters into the stationary equilibrium, or normal state. The Central Limit Theorem ensures that this state exists. See [Roberts & Rosenthal \(2004\)](#) for guaranteeing conditions and [Goodman & Weare \(2010\)](#) for *irreducibility* of the *emcee* stretch move. The stationary or normal state is reached when, in at least 95% of the iterations, the sample mean is bounded by two standard deviations of the sample, and the variance by the two standard deviation of the variance ¹³; see Fig. 3.6.

Once all parameters have entered the equilibrium state, I stop the *emcee* sampling once the criterium of [Gong & Flegal \(2016\)](#) ¹⁴ is fulfilled. We chose this criterium because it was developed for high-dimensional problems and tested on Hierarchical Bayesian Models, as in the present work. In this criterium, the MCMC chain stops once its effective sample size (ESS) is larger than a minimum sample size. This minimum is computed using the required accuracy, v , for each parameter confidence interval $(1 - \delta)100\%$. The ESS is the size that an independent and identically distributed sample must have to provide the desired accuracy on the parametric inference.

The *emcee* run stops once the ESS of the ensemble of walkers is greater than the minimum sample size needed for the required accuracy $\epsilon = 0.05$ on the 68% confidence interval ($\delta = 0.32$) of each parameter.

3.6 Codes

This Section sets out the details about the code I developed to perform the computation described throughout this chapter. First, I give a brief chronological description of the model development. Later, I will describe the details on the implementation of the charged PSO, the modified *emcee*, and the GMM used to describe the field population. Finally, I will end this Section detailing the hybrid high performance computing (Hybrid-HPC) code developed to minimise the computing time of the posterior distribution in the BHM.

The first version of the Bayesian Hierarchical Model was implemented by Ángel Berihuete in the package *Stan* ([Carpenter et al. 2017](#)). It comprised a Bayesian model

¹³ $sd(\sigma^2) = \sigma^2 \sqrt{\kappa/n + 2/(n-1)}$ with κ the kurtosis and n the sample size.

¹⁴ Implemented in the R package *mcmcse* ([Flegal et al. 2016](#))

of the ML model of Sarro et al. (2014). The proper motions were modelled using a single mixture of gaussians. The photometry was modelled with a Chebyshev polynomial parametrised by the length along the sequence. This length was found using a principal curve analysis and lacked physical interpretation.

I took this version and modified in the following aspects. I included the photometric and proper motions EMB sequence, the uncertainties both in proper motion and in photometry, and the width of the sequence modelled as the multivariate gaussian. Then, we realised that the principal curve analysis is not compatible with the deconvolution methodology. The principal curve analysis is driven by the elements with higher variance. Thus, we decided to parametrise the polynomials with the true colour, a more interpretable nuisance parameter that was later marginalised with the aid of a prior. For this prior we introduced the colour distribution as a GMM.

Previous to the introduction of the marginalisation of the nuisance parameters, the model worked fine on samples of a few hundreds of stars. Once the marginalisation was introduced, the computing time of the model increased dramatically, rendering its application to higher data sizes unpractical. At this point we decided to port the existing *Stan* code into *Python*¹⁵ so that we can work with the parallel *emcee* code. *emcee* proved to be of great use. Due to its parallelisation capabilities we were able to increase the data size from 2000 to 10,000 objects. Since the computing of the likelihood was the highest computational challenge, I developed my own routines to perform it on parallel. However, *CosmoHammer* (Akeret et al. 2013) turn out to be more efficient in distributing the parallel loads. I ported the BHM code into *CosmoHammer* and modified this last one. The modifications ranged from data files and log entries, to the introduction of priors and the handling of an Hybrid-HPC scheme using both MPI and multithreading. Despite the Hybrid-HPC scheme, the computing of the likelihood of a data set with 10^5 objects seemed unreachable. At this point, I performed two tasks, the first was to strip the code of all auxiliary libraries calls, and the second was the vectorisation of the majority of the operations. Since the parameters of the field were held fixed, the field likelihood was computed externally for each object. The code was then fed with the data, the field likelihood, and the auxiliary computations reduced at minimum. Among the reduced computation there are, for example, the Cholesky decompositions and matrix inversions of the covariance matrices of the uncertainty. Instead of doing these computation inside the code, the code was fed with the precomputed values. This further reduced the computing time.

Introducing PSO and later the charged PSO reduced even further the computing time. At this point the code was able to run on a data set with 10^5 objects. However, convergence of the MCMC still required several weeks of computations. Once the approximation to the marginalisation integral (Eqs. 3.29 and 3.30) was introduced, the computing time

¹⁵<https://www.python.org>

reduced far more. Finally, the tuning of the *emcee* parameters allowed us to increase the acceptance fraction, and reach convergence within for weeks of full computing time with a 80 cores computer cluster. It is indeed a very long time, however it is reasonable compared with our original estimates of approximately 2 years of computing time.¹⁶.

3.6.1 The modified charged PSO

As explained before, the charged PSO of [Blackwell & Bentley \(2002\)](#) was inappropriate to our objective. The metric of the parametric space of our problem is not isotropic, parameters have different length scales. For example, while fraction are constrained in the $[0, 1]$ interval, proper motions parameters are allowed in the range of proper motion measurements $[-99, 99] \text{ mas} \cdot \text{yr}^{-1}$. Therefore, the use of an isotropic metric results in a solution which is crowded in some parameters while is over dispersed in others. To solve this issue, I modified the charged PSO by measuring distance between particles and applying the electrostatic force independently on each parameter. In such a way, the electrostatic force plays a role only when the relative distance between particles is smaller than 10^{-10} . I found this value heuristically.

In the original version of [Blackwell & Bentley \(2002\)](#), each particle is subject to the acceleration,

$$\mathbf{a} = \sum_{i \neq j} \frac{q_i \cdot q_j}{r_{ij}^3} \cdot \mathbf{r}_{ij}, \quad p_{\text{core}} < r_{ij} < p \quad (3.42)$$

where q_i and q_j are the charges of particles i and j , r_{ij} is the distance between them. The distances p_{core} and p indicate the minimum and maximum distances at which the electrostatic force came into action. Outside this range, the electrostatic force is zero. In this equation, $\mathbf{r}_{ij} = \mathbf{x}_i - \mathbf{x}_j$, where $\mathbf{x}_i, \mathbf{x}_j$ are the positions of particles i and j . Also, $\mathbf{r}_{ij}, \mathbf{x}_i, \mathbf{x}_j \subset \mathbb{R}^d$, with d the dimension of the space.

In the modified version, the distance is measured independently in each dimension of the parametric space. Thus, $\mathbf{r}_{ij} = \{x_{1,i} - x_{1,j}, x_{2,i} - x_{2,j}, \dots, x_{d,i} - x_{d,j}\}$. Also the acceleration has the form,

$$\mathbf{a} = \sum_{i \neq j} \frac{q_i \cdot q_j}{r_{ij}^2} \cdot \mathbf{r}_{ij}, \quad 10^{-50} < \frac{r_{ij}}{r_{eq}} < \epsilon \quad (3.43)$$

and it is now applied over each dimension of the parametric space. The distance r_{eq} is that at which the velocity caused by the acceleration equals the mean velocity caused by the common PSO. ϵ is a free parameter, which as said previously was set heuristically to 10^{-10} .

¹⁶Today, the DANCe team is working on a GPU version of the code which is expected to compute the same amount of calculations in a couple of days.

3.6.2 Improvements of emcee

The modification I introduced in *emcee*, although very simple, improved the acceptance fraction and mixing of the particles. To allow the parallelisation, Foreman-Mackey et al. (2013) divide the ensemble of particles in two ensembles. In the original version, the particles in one ensemble use one and the same particle in complementary ensemble to compute their positions according to Eq. 3.39. In the modified versions, particles from one ensemble update their positions using a particle from the complementary ensemble, however, this particle is chosen randomly at each iteration.

In a private communication with David Foreman, the developer of *emcee*, he mentions that a similar modification was already introduced in a beta version of *emcee* code.

3.6.3 GMM for the field population

As mentioned earlier in this Chapter, the field population is modelled by means of two independent photometric and proper motion distributions. The MLE of the parameters of these distributions were found using the EM algorithm. The conventional EM algorithm for GMM (Dempster et al. 1977) for a mixture of M gaussians goes as follows. Given a set of parameters $\theta = \{w_i, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\}_{i=1}^M$, where w_i , $\boldsymbol{\mu}_i$, and $\boldsymbol{\Sigma}_i$ are the fraction, mean and covariance matrix of gaussian component i , the likelihood of the data is,

$$p(\{\mathbf{y}_n\}_{n=1}^N | \theta) = \prod_{n=1}^N \sum_{i=1}^M w_i \cdot \mathcal{N}(\mathbf{y}_n | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i). \quad (3.44)$$

To solve the problem, the EM algorithm requires a set of N variables, $\{\mathbf{z}_n\}_{n=1}^N$, of dimension M . The variable $z_{n,i}$ represent the probability that observation y_n was drawn from gaussian i . Therefore,

$$1 = \sum_{i=1}^M z_{n,i}. \quad (3.45)$$

These \mathbf{z} latent variables are found as

$$z_{n,i} = \frac{w_i \cdot \mathcal{N}(\mathbf{y}_n | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)}{\sum_{i=1}^M w_i \cdot \mathcal{N}(\mathbf{y}_n | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)}. \quad (3.46)$$

The EM works, as its name indicates, by maximising the expected value of the likelihood. This last is given by

$$E[p(\{\mathbf{y}_n\}_{n=1}^N | \theta)] = \prod_{n=1}^N \sum_{i=1}^M z_{n,i} \cdot w_i \cdot \mathcal{N}(\mathbf{y}_n | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i). \quad (3.47)$$

The previous expectation is maximal when,

$$w_i = \frac{1}{N} \sum_{n=1}^N z_{n,i}, \quad (3.48)$$

$$\boldsymbol{\mu}_i = \frac{1}{\sum_{n=1}^N z_{n,i}} \sum_{n=1}^N z_{n,i} \cdot \mathbf{y}_n, \quad (3.49)$$

$$\boldsymbol{\Sigma}_i = \frac{1}{\sum_{n=1}^N z_{n,i}} \sum_{n=1}^N z_{n,i} \cdot (\mathbf{y}_n - \boldsymbol{\mu}_i) \times (\mathbf{y}_n - \boldsymbol{\mu}_i)^T. \quad (3.50)$$

The modified version of the GMM, which includes a uniform distribution, is now a particular case of the GMM. This can be viewed as a gaussian distribution with fixed parameters and a constant probability c given by the uniform distribution. The new expectation is then,

$$E[p(\{\mathbf{y}_n\}_{n=1}^N | \theta)] = \prod_{n=1}^N \left[z_{n,0} \cdot w_0 \cdot c + \sum_{i=1}^M z_{n,i} \cdot w_i \cdot \mathcal{N}(\mathbf{y}_n | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \right]. \quad (3.51)$$

The maximisation step remains identical except for the indices. There are now $M + 1$ fractions w_i , with $i = 0, 1, \dots, M$, and the means and covariances run from $i = 1, \dots, M$.

Regarding the photometric GMM of the field. Since photometry has missing values, we use the EM algorithm of [McMichael \(1996\)](#). This algorithm was developed to obtain the MLE of data with missing values. A more recent version has been developed by [Lin et al. \(2006\)](#), however, this assumes that the missing data is randomly distributed (missing at random). However, as mentioned earlier, the missing values in the photometry of the DANCE DR2 are not randomly distributed.

In the algorithm of [McMichael \(1996\)](#), there is a set of N gain matrices, one for each datum. Each M_n matrix is an identity matrix in which the rows of the corresponding missing value have been deleted. Thus, the expected value of the likelihood is now,

$$E[p(\{\mathbf{y}_n\}_{n=1}^N | \theta)] = \prod_{n=1}^N \sum_{i=1}^M z_{n,i} \cdot w_i \cdot \mathcal{N}(\mathbf{y}_n | M_n \boldsymbol{\mu}_i, M_n \boldsymbol{\Sigma}_i M_n). \quad (3.52)$$

The maximisation step is now,

$$w_i = \frac{1}{N} \sum_{n=1}^N z_{n,i}, \quad (3.53)$$

$$\boldsymbol{\mu}_i = \frac{\sum_{n=1}^N z_{n,i} \cdot H_n \mathbf{y}_n}{\sum_{n=1}^N z_{n,i} H_n M_n} \quad (3.54)$$

with

$$H_i = M_i^T (M \boldsymbol{\Sigma}_i M^T)^{-1} \quad (3.55)$$

The maximisation step has no analytical solution for the covariance matrix, therefore, a modified steepest descent is used

$$\Sigma_i \leftarrow \Sigma_i + \frac{\rho}{2} \cdot \Sigma_i \Delta_i \Sigma_i, \quad (3.56)$$

with and Δ_i given by

$$\Delta_i = \frac{1}{\sum_{n=1}^N z_{n,i}} \cdot \sum_{n=1}^N z_{n,i} \cdot [H_n(\mathbf{y}_n - M_n \boldsymbol{\mu}_i) \times (\mathbf{y}_n - M_n \boldsymbol{\mu}_i)^T H_n^T - H_n I]. \quad (3.57)$$

This algorithm preserves the monotonic convergence of the conventional EM, and returns positive definite matrices provided that $\rho < 2$. For details and validation of this algorithm see [McMichael \(1996\)](#).

3.6.4 Hybrid-HPC implementations

As I outlined before, the parallel computing approach was an unavoidable step. Once the code was ported to *Python* and *CosmoHammer*, I modified this last one to better fit our needs. The modifications were mainly on the management of input and output files and the python *multiprocessing* package for the multithreading computing of likelihood. Also, I striped some of its original functions to reduce memory usage and implemented some others, like the use of initial positions for the particles.

In the Hybrid-HPC approach, the particles of *emcee* are distributed on the nodes of the computing cluster by means of the MPI protocol. Then, each core in the node computes the likelihood of one fraction of the objects in the data set. This Hybrid-HPC code was implemented and tested in different computing cluster architectures. For the cluster at the Centre of Astrobiology (CAB, Villanueva de la cañada, Madrid, Spain), I used a configuration of 6 nodes each with 12 cores. For the cluster at the University of Cádiz, (Andalucía, Spain) I used a configuration of 5 nodes each with 16 cores.

However, the Hybrid-HPC approach was not the best solution at the CIMENT clusters of the University of Grenoble Alpes. At this cluster (I used Froggy), the code continuously render errors of communication. For this reason, I implemented a only-MPI version of the code. In this version, the multithreading approach is left aside. Instead, the totality of the available cores is fully dedicated to the computing of the likelihood. Each of the n cores computes the likelihood of one n th of the objects in the data set. Once the likelihood of all particles has been computed, the master node evaluates the new positions of the particles.

I finish this chapter with a brief description of the difficulties faced on the development and testing of the BHM code. As the code evolved in complexity, my computational skills were compelled to evolve as well. I started by learning R and solving some toy problems on it. Later, when the dimensionality of the posterior increased I learned *Stan*. When

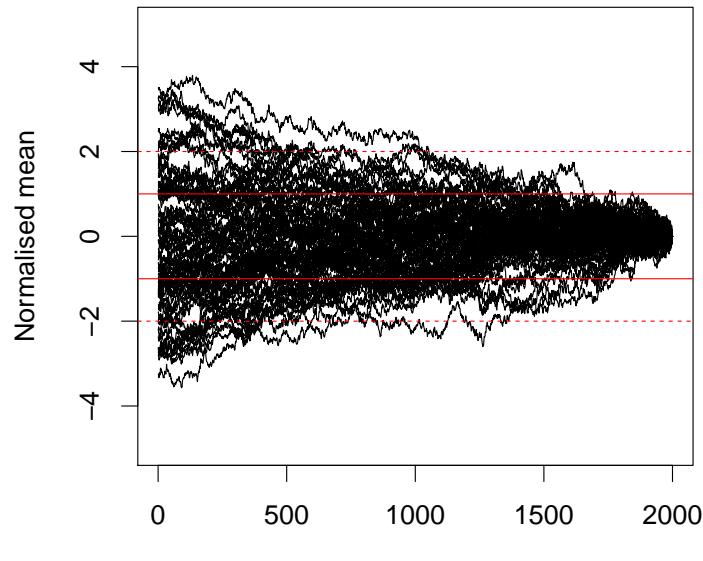
the data set increased in size, we faced the parallelisation, so I learned Python and MPI. Once these versions were operable, I was forced to deal with libraries, from the common, numpy, scipy and numba, to the linking of modules and libraries. Finally, when dealing with several clusters I learned the queue languages Condor, slurm and OAR. Currently, I am working on the improvement (memory allocation and data distribution) of GPU implementation of the BHM code.

Table 3.1: Parameters names, symbols, and priors.

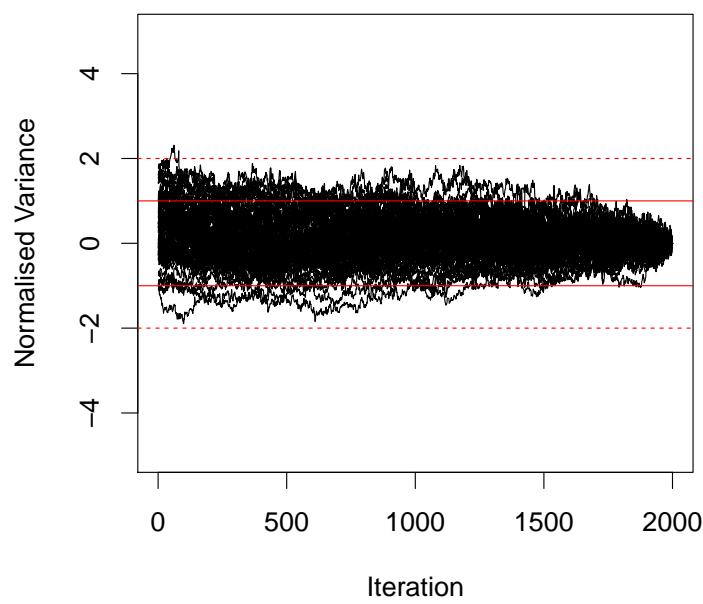
Name	Symbol	Prior	Name	Symbol	Prior
Field fraction	π	Dirichlet(α)	Coefficient [1,1]	$\beta_{Y,1}$	Normal(μ_β, σ_β)
Cs fraction	π_{CB}	Dirichlet(α_{Cs})	Coefficient [1,2]	$\beta_{Y,2}$	Normal(μ_β, σ_β)
Cs PM fraction 1	$\pi_{Cs,1}$	Dirichlet(α_{Cs})	Coefficient [1,3]	$\beta_{Y,3}$	Normal(μ_β, σ_β)
Cs PM fraction 2	$\pi_{Cs,2}$	Dirichlet(α_{Cs})	Coefficient [1,4]	$\beta_{Y,4}$	Normal(μ_β, σ_β)
Cs PM fraction 3	$\pi_{Cs,3}$	Dirichlet(α_{Cs})	Coefficient [1,5]	$\beta_{Y,5}$	Normal(μ_β, σ_β)
Bs PM fraction 1	$\pi_{Bs,1}$	Dirichlet(α_{Bs})	Coefficient [1,6]	$\beta_{Y,6}$	Normal(μ_β, σ_β)
Color fraction 1	$\pi_{CI,1}$	Dirichlet(α_{CI})	Coefficient [1,7]	$\beta_{Y,7}$	Normal(μ_β, σ_β)
Color fraction 2	$\pi_{CI,2}$	Dirichlet(α_{CI})	Coefficient [2,1]	$\beta_{J,1}$	Normal(μ_β, σ_β)
Color fraction 3	$\pi_{CI,3}$	Dirichlet(α_{CI})	Coefficient [2,2]	$\beta_{J,2}$	Normal(μ_β, σ_β)
Color fraction 4	$\pi_{CI,4}$	Dirichlet(α_{CI})	Coefficient [2,3]	$\beta_{J,3}$	Normal(μ_β, σ_β)
Mean color 1	$\mu_{CI,1}$	Unifrom(rg_{CI})	Coefficient [2,4]	$\beta_{J,4}$	Normal(μ_β, σ_β)
Mean color 2	$\mu_{CI,2}$	Unifrom(rg_{CI})	Coefficient [2,5]	$\beta_{J,5}$	Normal(μ_β, σ_β)
Mean color 3	$\mu_{CI,3}$	Unifrom(rg_{CI})	Coefficient [2,6]	$\beta_{J,6}$	Normal(μ_β, σ_β)
Mean color 4	$\mu_{CI,4}$	Unifrom(rg_{CI})	Coefficient [2,7]	$\beta_{J,7}$	Normal(μ_β, σ_β)
Mean color 5	$\mu_{CI,5}$	Unifrom(rg_{CI})	Coefficient [3,1]	$\beta_{H,1}$	Normal(μ_β, σ_β)
Variance color 1	$\sigma_{CI,1}$	HalfCauchy($0, \eta$)	Coefficient [3,2]	$\beta_{H,2}$	Normal(μ_β, σ_β)
Variance color 2	$\sigma_{CI,2}$	HalfCauchy($0, \eta$)	Coefficient [3,3]	$\beta_{H,3}$	Normal(μ_β, σ_β)
Variance color 3	$\sigma_{CI,3}$	HalfCauchy($0, \eta$)	Coefficient [3,4]	$\beta_{H,4}$	Normal(μ_β, σ_β)
Variance color 4	$\sigma_{CI,4}$	HalfCauchy($0, \eta$)	Coefficient [3,5]	$\beta_{H,5}$	Normal(μ_β, σ_β)
Variance color 5	$\sigma_{CI,5}$	HalfCauchy($0, \eta$)	Coefficient [3,6]	$\beta_{H,6}$	Normal(μ_β, σ_β)
Mean PM Cs[1,1]	$\mu_{Cs,1}$	Normal($\mu_{\mu_{pm}}, \Sigma_{\mu_{pm}}$)	Coefficient [3,7]	$\beta_{H,7}$	Normal(μ_β, σ_β)
Mean PM Cs[1,2]	$\mu_{Cs,2}$	Normal($\mu_{\mu_{pm}}, \Sigma_{\mu_{pm}}$)	Coefficient [4,1]	$\beta_{K,1}$	Normal(μ_β, σ_β)
Variance Cs[1,1]	$\Sigma_{Cs,1,1}$	Half-t(ν, A_{pm})	Coefficient [4,2]	$\beta_{K,2}$	Normal(μ_β, σ_β)
Variance Cs[1,2]	$\Sigma_{Cs,1,2}$	Half-t(ν, A_{pm})	Coefficient [4,3]	$\beta_{K,3}$	Normal(μ_β, σ_β)
Variance Cs[1,3]	$\Sigma_{Cs,1,3}$	Half-t(ν, A_{pm})	Coefficient [4,4]	$\beta_{K,4}$	Normal(μ_β, σ_β)
Variance Cs[2,1]	$\Sigma_{Cs,2,1}$	Half-t(ν, A_{pm})	Coefficient [4,5]	$\beta_{K,5}$	Normal(μ_β, σ_β)
Variance Cs[2,2]	$\Sigma_{Cs,2,2}$	Half-t(ν, A_{pm})	Coefficient [4,6]	$\beta_{K,6}$	Normal(μ_β, σ_β)
Variance Cs[2,3]	$\Sigma_{Cs,2,3}$	Half-t(ν, A_{pm})	Coefficient [4,7]	$\beta_{K,7}$	Normal(μ_β, σ_β)
Variance Cs[3,1]	$\Sigma_{Cs,3,1}$	Half-t(ν, A_{pm})	Covariance Phot[1]	$\Sigma_{clus}[1]$	Half-t(ν, A_{ph})
Variance Cs[3,2]	$\Sigma_{Cs,3,2}$	Half-t(ν, A_{pm})	Covariance Phot[2]	$\Sigma_{clus}[2]$	Half-t(ν, A_{ph})
Variance Cs[3,3]	$\Sigma_{Cs,3,3}$	Half-t(ν, A_{pm})	Covariance Phot[3]	$\Sigma_{clus}[3]$	Half-t(ν, A_{ph})
Variance Cs[4,1]	$\Sigma_{Cs,4,1}$	Half-t(ν, A_{pm})	Covariance Phot[4]	$\Sigma_{clus}[4]$	Half-t(ν, A_{ph})
Variance Cs[4,2]	$\Sigma_{Cs,4,2}$	Half-t(ν, A_{pm})	Covariance Phot[5]	$\Sigma_{clus}[5]$	Half-t(ν, A_{ph})
Variance Cs[4,3]	$\Sigma_{Cs,4,3}$	Half-t(ν, A_{pm})	Covariance Phot[6]	$\Sigma_{clus}[6]$	Half-t(ν, A_{ph})
Mean PM Bs[1,1]	$\mu_{Bs,1}$	Normal($\mu_{\mu_{pm}}, \Sigma_{\mu_{pm}}$)	Covariance Phot[7]	$\Sigma_{clus}[7]$	Half-t(ν, A_{ph})
Mean PM Bs[1,2]	$\mu_{Bs,2}$	Normal($\mu_{\mu_{pm}}, \Sigma_{\mu_{pm}}$)	Covariance Phot[8]	$\Sigma_{clus}[8]$	Half-t(ν, A_{ph})
Variance Bs[1,1]	$\Sigma_{Bs,1,1}$	Half-t(ν, A_{pm})	Covariance Phot[9]	$\Sigma_{clus}[9]$	Half-t(ν, A_{ph})
Variance Bs[1,2]	$\Sigma_{Bs,1,2}$	Half-t(ν, A_{pm})	Covariance Phot[10]	$\Sigma_{clus}[10]$	Half-t(ν, A_{ph})
Variance Bs[1,3]	$\Sigma_{Bs,1,3}$	Half-t(ν, A_{pm})	Covariance Phot[11]	$\Sigma_{clus}[11]$	Half-t(ν, A_{ph})
Variance Bs[2,1]	$\Sigma_{Bs,2,1}$	Half-t(ν, A_{pm})	Covariance Phot[12]	$\Sigma_{clus}[12]$	Half-t(ν, A_{ph})
Variance Bs[2,2]	$\Sigma_{Bs,2,2}$	Half-t(ν, A_{pm})	Covariance Phot[13]	$\Sigma_{clus}[13]$	Half-t(ν, A_{ph})
Variance Bs[2,3]	$\Sigma_{Bs,2,3}$	Half-t(ν, A_{pm})	Covariance Phot[14]	$\Sigma_{clus}[14]$	Half-t(ν, A_{ph})
			Covariance Phot[15]	$\Sigma_{clus}[15]$	Half-t(ν, A_{ph})

Table 3.2: Hyper-parameters for different blocks of the Hierarchical Bayesian Model. The upper and middle blocks correspond, respectively, to hyper-parameters of the proper motions and photometric models. The lower block to hyper-parameters shared by both models.

Hp.	Value
α_{Cs}	{1, 1, 5, 5}
α_{Bs}	{1.2, 8.8}
A_{pm}	{ 10^5 , 10^5 }
$\mu_{\mu_{pm}}$	{16.30, -39.62}
$\Sigma_{\mu_{pm}}$	{36.84, 1.18, 40.71}
α_{CI}	{1, 1, 1, 1, 1}
rg_{CI}	{0.8, 8}
η	100
μ_{β_Y}	{7.65, 11.47, 10.66, 16.33, 16.49, 21.44, 22.49}
μ_{β_J}	{7.61, 11.52, 10.20, 15.66, 15.58, 19.88, 21.16}
μ_{β_H}	{7.63, 10.88, 9.50, 15.19, 15.04, 18.68, 20.64}
μ_{β_K}	{7.55, 10.81, 9.32, 14.79, 14.62, 17.63, 20.24}
σ_β	{1, 1, 1, 1, 0.5, 0.1}
A_{ph}	{10, 10, 10, 10, 10}
α	{9.8, 0.2}
α_{CB}	{8, 2}
ν	3



(a)



(b)

Figure 3.6: Normalised mean (a) and variance (b) of each parameter in BHM model as functions of iterations. The normalisation values are the mean and variance of the ensemble of particles positions at the last iteration. Red lines show one and two sigma levels of these normalisation values. Only shown the last 2000 iterations previous to stopping the algorithm.

Chapter 4

Results

In this Chapter I characterise the methodology detailed in Chapter 3 and then I applied it to the Pleiades DANCe DR2 data set (Sect. 2.7). To characterise the methodology as a classifier, I measure its precision and accuracy when applied on synthetic data where the true members of the cluster are known. With this characterisation, I am able to obtain an optimal probability threshold, but only for classification purposes. Afterwards, I apply the methodology to the Pleiades DANCe RDR2 and I find the candidate members of the cluster using the optimal probability threshold. Then, I compare these candidate members with those found by previous studies.

Later, I analyse the main results of this work, those that fulfil the objective: the statistical distributions that characterise of the cluster population. Then, I give the details of the spatial, velocity, luminosity and mass distributions. Finally, I end this Chapter describing the physical scenario of the evolution of the mass distribution of the Pleiades by comparing it with other mass distribution of younger and older clusters.

4.1 Performance of the classifier

As mentioned earlier, the main objective of the methodology of the BHM is the statistical characterisation of the NYOC populations. However, as a by product, it also obtains the individual membership probability distributions of the objects comprising the data set. These membership probability distributions, together with a probability threshold, allow a direct classification of the objects into cluster and field members. The classification resulting from this procedure, as any other measured property, has an uncertainty. By evaluating this uncertainty under the results of synthetic data (in which the true members are known) we are able to measure the accuracy and precision of the classification process as a function of the probability threshold used. This section explains how an objective probability threshold can be found by maximising the accuracy of the classifier.

To measure the accuracy of our classifier, I test it over synthetic data sets that resemble

the real RDR2 (Section 2.7.2). An ideal test to our classifier will be to apply it over well known dataset in which tags of cluster and field members were already present. However, if we may have access to these tags, a classifier may not be needed. The Pleiades cluster being one of the most studied cluster in history, it is the NYOC with most of these tags (see Section 4.2). This is one the reasons for which we decided to benchmark our methodology on it. In spite of the large number of candidate members for the Pleiades clusters, the synthetic data and its true tags are still needed. The reasons are the following. First, the list of candidate members provided by the literature is not infallible. We can never be sure that this list is complete and unpolluted. Here it is important to note that the astrophysical domain of the phenomenon marks a very important distinction with common supervised classification methods. A perfect and real training set is not available. It must be created from simulations. Second, even if the most probable candidate members from the literature were used as a training set (as done for example by Sarro et al. 2014) the very faint end of the magnitude distribution (represented by brown-dwarfs) is still a *terra incognita* where candidate members are scarce or not even exist.

Thus, to overcome the problem of the true tags, we decided to create synthetic data sets. These synthetic true tags, and therefore, the results obtained from them relay under the assumption that our cluster and field models resemble the real data. I am aware that these models are far from perfect, but so far this assumption provides the best option. Although this assumption enable us to quantify the internal consistence (precision and accuracy) of our classifier, it does not give any indication about possible biases in the model. To explore this possibility, in the next Section, I compare our real data classification results with those of the literature.

The random nature of the synthetic data sets demands the repetition of the data sets and results. This repetitions avoid any bias caused by excursions of the random number generator (bad luck), and more importantly, they allow us to compute the uncertainty in the accuracy and precision of the classifier. As explained before, the methodology presented in this work is computing demanding. Thus, to be able to repeat at least five times the results of the synthetic data sets, we further reduce the data set size. The inference process in a data set of 10^4 objects demands almost week of computing time. Thus, we decided that such data set provided a good compromise between computing time and number of objects. Furthermore, to provide a better estimate of the contamination rate on the results over the real RDR2 data set (the one with the 10^5 objects) we chose to work with the 10^4 objects with high membership probability according to Bouy et al. (2015). Since this sample is relatively more entangled with the cluster than that of the RDR2, we assume that the contamination rate we measure on this sample will be comparable or even higher than that hypothetically obtained over the 10^5 RDR2.

Briefly, to create the synthetic data set, the procedure is the following. First, using the methodology of the previous Chapter, I obtain a sample of the posterior distribution of

the parameters in the model given the 10^4 real data set. Then, I chose the particle with highest posterior probability as the MAP estimate of the posterior distribution. Using this particle positions in the parametric space, I generate five synthetic data sets of 10^4 objects each. Then, I tag these objects according to their parent population: cluster or field. Afterwards, using the vector of synthetic values of each object, I assign their uncertainties and missing value entries (more details below).

Finally, I run the methodology over the five synthetic data sets, and obtain a sample of the individual membership probability distributions of each synthetic object. Then, I compare the true tags with the measured ones as function of the probability threshold.

To further test the performance of the classifier, I apply it on a synthetic data set in two different cases. In the first case the data set is one of the five synthetic ones. Thus it contains objects with missing value entries in their vector of observables. In the second case the data set is the same as in the previous case, but this time all the objects have fully observed vectors (i.e. do not have missing value entries). The comparison of the results rendered by these two cases allows us to quantify the impact of missing values.

Objects in the Pleiades DANCE DR2 have full observed proper motions vectors. Since missing values appear only in the photometric vectors, in synthetic data sets I only include missing values in the photometry. Only $\sim 1\%$ of the objects in the RDR2 have full observed vectors (i.e. no missing value entries). Furthermore, the distribution of missing entries is not random and depends on the magnitudes of the objects (see Fig. 3.2). Therefore, to better reproduce this distribution in the synthetic data sets, for each synthetic datum, I use the mask of missing entries of one of its closer neighbours in the real data set. Here, distance is measured in the euclidean sense. If I were to use the missing value mask of the nearest neighbour in the real data set, then I will obtain a biased sample in which objects with full observed vectors (with no missing entries) will be underestimated. This is the inevitable consequence of the fact that euclidean distances measured in subspaces, which result from removing the dimensions of the missing entries, are small, or at most equal, than those measured in the complete space.

To reproduce the mask of missing entries, I chose from among those masks of the closer neighbours in the available CMDs: $\{K_s, J - K_s\}$, $\{J, J - H\}$, $\{K_s, H - K_s\}$, $\{J, Y - J\}$, $\{K_s, i - K_s\}$. These CMD are formed with the bands and colours of the magnitudes with fewer missing values, in decreasing order (see Table 2.1). The missing entries mask to reproduce in the synthetic objects is chosen as follows. First, for each CMD subspace I find the set of objects, from the real data set, with fully observed entries, I call it $C_{or,i}$ and its fraction from the total, f_r . Then, I take a random sample from the synthetic data whose fraction, f_s equals f_r . For objects in this sample I assign the missing value pattern of the nearest neighbour in $C_{or,i}$. I repeat this procedure for the rest of the CMDs. In this way, the synthetic data sets have fractions of objects with and without missing entries, similar to those of the objects in the real data set.

To assign uncertainties I proceed as follows. For uncertainties in the proper motions I use those of the nearest neighbour from the real data set. If I were to use the nearest neighbour scheme for uncertainties in the photometry, then I will obtain biased results. These uncertainties will be biased towards those of the less precise measurements. Again, this is a consequence of the missing entries in the observable vectors. The euclidean metric results in the preferential choosing of objects with missing values. These missing values occur mostly at the faint end, where uncertainties are larger. Therefore, these uncertainties will be biased towards these larger values. To avoid this issue, I fit 8th degree Chebyshev polynomials to the uncertainties as a function of the magnitudes. Then, I use these polynomials to establish the photometric uncertainties of the synthetic data sets.

Once the synthetic data sets were created I run the methodology on each of them and recover the membership probability distribution of each object. Then, I classify each object as cluster or field member. To classify an object as a cluster member, the mode of its cluster membership probability distribution must be higher than the probability threshold. Otherwise it is classified as a field object.

The classifier performance was measured by counting the true positives (TP, cluster members correctly classified), true negatives (TN, field members correctly classified), false positives (FP, field members classified as cluster members) and false negatives (FN, cluster members classified as field members) recoveries as a function of the probability threshold. With them I calculate the following quantities: the true positive rate (TPR), which is the ratio of true positives over the sum of true positives plus false negatives, the contamination rate (CR), which is the ratio of false positives over the sum of false positives plus true positives, the precision or positive predictive value (PPV), which is the ratio of true positives over the sum of true positives plus false positives, and, the accuracy (ACC), which is the ratio of the sum of true positives plus true negative over the sum of true and false positives and negatives.

These are,

$$\begin{aligned} TPR &= \frac{TP}{TP + FN} \\ CR &= \frac{FP}{FP + TP} \\ PPV &= \frac{TP}{TP + FP} \\ ACC &= \frac{TP + TN}{TN + FN + TP + FP}, \end{aligned}$$

which are defined at each probability threshold. I use the results of the five synthetic data sets to quantify the uncertainties of the previous quantities.

In Fig. 4.1, I show the mean and uncertainty of the TPR and CR measured on the five synthetic data sets. The uncertainty is represented by the maximum deviation from the mean. Also, this Figure shows the TPR and CR measured on a synthetic data set with

fully observed objects (i.e. objects with non missing entries). As this Figure shows, the missing values have a negative impact in our classification. They diminish the TPR and increase the CR. This negative impact is expected since the observables we are using are highly discriminant in the classification process (see Section 2.7.1). Since cluster and field are highly entangled in the 10^4 objects synthetic samples, when one of these observables is missing the classification is more uncertain or it could even be biased. Interestingly, the CR above probability threshold 0.8 is independent of the missing values and remains low ($\lesssim 5\%$). In spite of the negative impact of missing values, the methodology delivers low contamination rates ($\lesssim 5 - 10\%$) and high recovery rates ($\lesssim 90 - 96\%$) for probability thresholds in the $0.5 - 0.9$ range.

In Figure 4.1, I also show the CR and TPR of Sarro et al. (2014) (reported in their Table 4). Previous to discuss the differences between both works I inform the reader about the unfairness of this comparison. First, in the works of Sarro et al. (2014) and Bouy et al. (2015), the generative models are constructed using only fully observed objects (i.e. without missing entries), these objects represent only $\sim 1\%$ of our RDR2. Afterwards, they apply those models to all the objects in the DANCE DR2 data set (i.e. objects with and without missing entries). Thus, their results are more similar to those I find on the synthetic data set with only fully observed objects (blue lines in Fig. 4.1). Second, the synthetic data sets in which both works measure the TPR and CR are different. They are constructed with different generative models, different number of objects, and different missing value distributions.

From the comparison between Sarro et al. (2014) TPR and CR, with the ones I find on the synthetic data set comprising only fully observed objects, we see the following. The TPR of both works agree within the uncertainties. The CR of both works agree, within the reported uncertainties, for probability thresholds below 0.8. At higher probability thresholds, our methodology delivers lower CR. Even in the case of the unfair comparison between the TPR and CR of Sarro et al. (2014) with those I measure on the synthetic data sets including objects with missing entries, our CR outperforms that reported by Sarro et al. (2014) at the small price of a lower ($\approx 4\%$) TPR.

Now, I describe the procedure to set an optimal probability threshold. This probability threshold, although not needed to obtain the posterior distribution of the parameters modelling the cluster population, is needed, however, to objectively classify an object as a cluster member. I establish this threshold using only the synthetic data sets containing objects with missing entries in their vector of measurements. The approach I use to set this probability threshold is that of the maximum accuracy (ACC) for the classification.

Figure 4.2 shows the ACC and the PPV of the classifier when it is applied on synthetic data sets containing objects with missing value entries. The lines and the grey regions depict, respectively, the mean and the maximum deviations of the results of the five synthetic data sets. The maximum deviation is used as a proxy for the uncertainty.

The highest mean accuracy, $\text{ACC}=96.5 \pm 0.1\%$, happens at probability threshold $p_t = 0.84$. Thus, I chose it as the optimal classification threshold. At this value, the CR is $4.3 \pm 0.2\%$, the TPR is $90.0 \pm 0.05\%$, and the PPV is $95.6 \pm 0.2\%$.

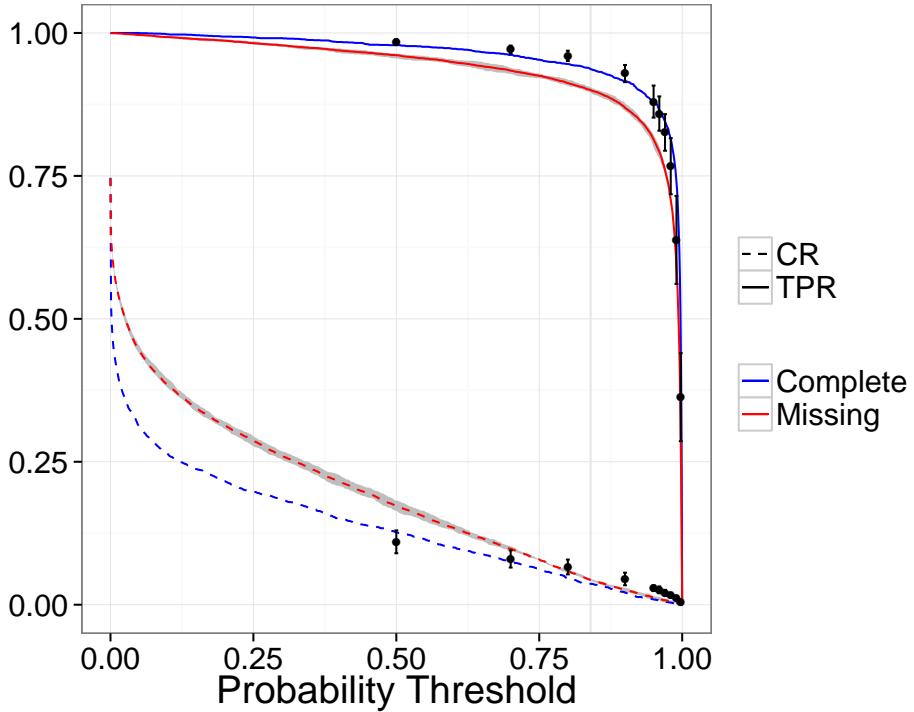


Figure 4.1: The mean TPR (solid line) and CR (dashed line) resulting from five synthetic data sets including objects with missing entries (red lines). Also the TPR and CR resulting from a synthetic data set comprising only objects with fully observed vectors (blue lines). The shaded regions (grey) show the uncertainties computed from the five synthetic data sets. The black dots show the TPR and CR reported by [Sarro et al. \(2014\)](#) for their model. See text for warnings on this comparison. Reproduced from Figure 3 of [Olivares et al. \(2017\)](#), *The Seven sisters DANCe III: Hierarchical Bayesian model*, A&A, submitted.

We investigate further on the impact that objects with missing value entries in their observations vectors have on our methodology. Particularly, in the following I analyse possible biases introduced by these objects. To do this, I compare the membership probabilities, summarised by the mode, recovered after inferring the model on two synthetic data sets of. These two data sets are identical except that in one of them some entries in the vector of observables were masked as missing (using the procedure previously described).

In Fig. 4.3, I compare the mode of the membership probabilities. The horizontal axis

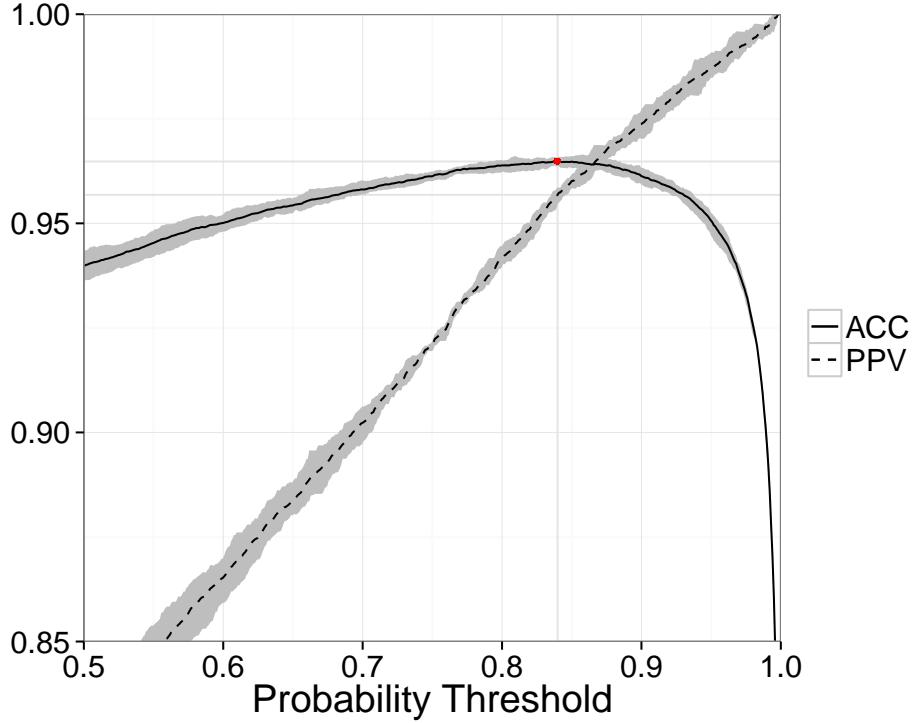


Figure 4.2: Mean accuracy (ACC, solid line) and precision (PPV, dashed line) of the classifier as a function of probability threshold. The shaded regions shows the uncertainties computed from the five synthetic data sets. The higher accuracy is obtained at $p_t = 0.84$ (red dot). Reproduced from Figure 3 of Olivares et al. (2017), *The Seven sisters DANCE III: Hierarchical Bayesian model*, A&A, submitted.

shows the membership probabilities of the data set with fully observed objects (I call this case the complete one). The vertical axis shows the membership probabilities of the same objects but in which some entries were masked as missing (I call this case the Incomplete one). As can be seen in this Fig., the missing values impact our results by spreading the membership probabilities. Ideally, we would like to recover membership probabilities following the line of slope one. This is the case of some fully observed objects (red squares) in the data set containing objects with missing entries. The most striking deviations come from those objects with the *CI* masked as missing (enclosed in black). The BHM methodology uses the *true CI* to prescribe the *true photometry*. Also, it uses the observed *CI* to constrain the marginalisation integral of the *true CI*. Thus, as expected, a missing *CI* produces a spread in the membership probability.

The objects with a missing *CI* show two different behaviours. In one case, there are objects with membership probabilities from the fully observed data set (horizontal axis)

that have an overestimated membership probability in the missing entries data set (vertical axis). In the other case, there are objects that have underestimated probabilities in the missing entries data set. These objects correspond to those seen in the combed area below the line of unit slope.

Objects in the first case increase the CR, and their effect can be seen by the difference between red and blue dashed lines in Fig. 4.1. On the other hand, the objects in the second case diminish the TPR, their effect can also be seen by the difference between red and blue solid lines in Fig. 4.1. The increase in CR reaches its maximum near probability zero in the horizontal axis (Complete case) and goes to zero at probability thresholds of ≈ 0.9 . Therefore, the impact this increased CR has in our results is marginal. For example, at the optimal probability threshold $p_t = 0.84$, the increase of CR due to objects with missing entries represent only 1.8%. This correspond to the objects in box region of Fig. 4.3 (upper left corner). However, the objects in the second case, those that diminish the TPR, represent the typical unavoidable loss of members due to their missing entries. These amount to a 4% loss in the TPR, at the optimal probability threshold, $p_t = 0.84$.

The bias introduced in the recovered membership probabilities due to objects with missing value entries, can be quantified using the root-mean-square (rms) of the difference between the means of the two recovered membership probabilities (Complete and Incomplete cases). The total rms is 0.12. On the one hand, fully observed objects in both data sets (Complete and Incomplete cases) have a rms of only 0.02. On the other hand, objects with missing entries, excluding those with missing CI , have a rms of 0.08. The rms of objects lacking the CI is 0.14. The previous effects show an overall agreement between results on data sets with and without objects with missing entries. Nonetheless, care must be taken when dealing with individual membership probabilities. An object with a missing value in the Y, J, H and K_s may have a diminished membership probability (with a rms of 0.08), while an object with a missing CI may show an increased membership probability (with a rms of 0.14).

However, as have been mentioned before, the methodology described in this work aims at the statistical distributions of the cluster population. The individual membership probabilities are just a useful by product. The methodology develop here, though, works by ensuring that each object contributes to the posterior distribution of the parameters modelling the cluster population, proportionally to its cluster membership probability. In this sense our results are free of any possible bias introduced by cuts in the membership probability. Nevertheless, there is still contamination, particularly from objects with missing entries, as seen before. This contamination in the statistical distributions that we aim to obtain must be quantified. To do this, I compute the expected value of the CR found in this section. It is $\langle CR \rangle = 5.8 \pm 0.2\%$. In this expected value, each CR contributes proportionally to the probability threshold at which it is measured. Since the vast contribution to this CR comes from probability thresholds below 0.2 (see Fig. 4.1),

the expected value of the CR remains low.

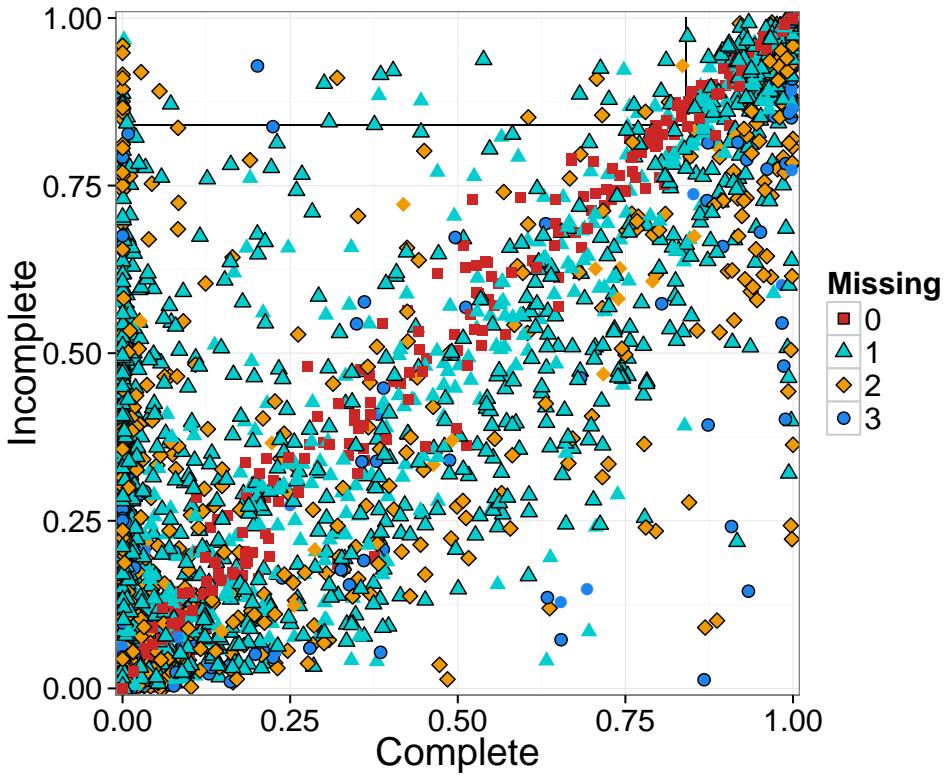


Figure 4.3: Comparison between the cluster membership probabilities recovered from the synthetic data set with objects having missing value entries (vertical axis, labeled Incomplete), and, the synthetic data set with fully observed objects (horizontal axis, labeled Complete). The colour and shape indicate the amount of missing entries. The symbols enclosed in black indicate a missing CI . The top left box contains objects considered as contaminants due to missing values at the probability threshold $p_t = 0.84$. Reproduced from Figure 4 of Olivares et al. (2017), *The Seven sisters DANCe III: Hierarchical Bayesian model*, A&A, submitted.

In statistical science, particularly in machine learning, it is sometimes useful to analyse the performance of a binary classifier by the receiver operating characteristic curve, the ROC curve. It plots a visual diagnostic of the ability of a classifier to perform its job. The ROC curve plots the TPR as a function of the FPR for all possible values of the probability threshold. A perfect classifier would be that in which the TPR=1 and the FPR=0 for some probability threshold. On the other hand, a random classifier would be that with TPR=FPR at all probability thresholds. Such classifier has a line of slope one in its ROC curve. Furthermore, the quantitative diagnostic for a binary classifier is the area under the ROC curve (AUC). As its name indicates, the AUC is the integral of the

ROC curve. Thus a random classifier has a AUC of one half, while a perfect classifier has a AUC=1. In Fig. 4.4, I show the ROC curve for our classifier when applied over synthetic data containing objects with missing entries. It is the ROC of one of the five synthetic realisations described throughout this section. As can be seen from this Figure our classifier does an excellent job, with an AUC=0.992.

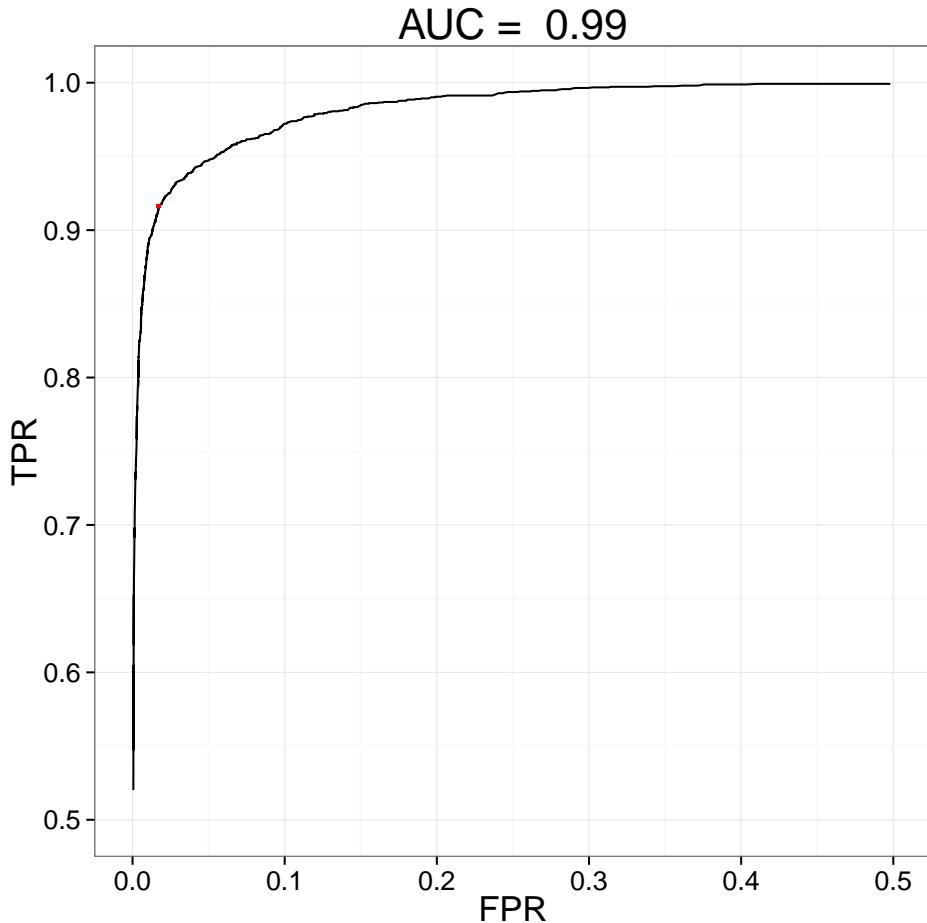


Figure 4.4: ROC curve of the BHM by-product classifier when applied on the synthetic data set containing objects with missing entries. As can be seen, the AUC=0.992 diagnose it as an excellent classifier.

4.2 Comparison with the literature

In Chapter 2, I mentioned that the most important works on the Pleiades members are those of Stauffer et al. (2007); Lodieu et al. (2012); Sarro et al. (2014); Bouy et al. (2015). In this section I will compare the list of candidate I members found with the most recent study of the Pleiades members, that of Bouy et al. (2015). Later, for the sake of completeness, I will compare my list of candidate members with the compilation made by Stauffer et al. (2007).

IS ALL THIS SECTIONING REQUIRED.

REJECTED CANDIDATES? OF WHOM? BY WHOM? AVOID CONFUSSION.

4.2.1 Candidate members from Stauffer et al. (2007)

REFER AS LIST I AND LIST II AVOID CONFUSION.

Stauffer et al. (2007) published two list of candidate members. The first one contains 1417 objects compiled from the literature (their Table 2). These objects were classified as candidate members by several authors along modern astronomy. As Stauffer et al. (2007) mention, this list is inhomogeneous, incomplete and certainly includes non members. Their second list contains 55 candidate members (their Table 5) found using infrared photometry, and proper motions.

Cross matching these two lists with the candidate members of Bouy et al. (2015), shows that only 1288 of the 1417 objects on Stauffer et al. (2007) Table 2 were classified by Bouy et al. (2015) as candidate members. From these, 1132 objects come from the DANce survey and the rest from Tycho+DANCe, see Appendix B of Bouy et al. (2015). Also, only 34 of the 55 new candidate members of Stauffer et al. (2007) were classified by Bouy et al. (2015) as candidate members. In Figures 4.5 - 4.6 I show the candidate members of Stauffer et al. (2007), and Bouy et al. (2015). As mentioned by Stauffer et al. (2007), many of the candidate members in their Table 2 are most probably non members. These objects lay far from the locus of the rest of the candidates. Similarly, the majority of the rejected candidates of Stauffer et al. (2007), 15 of those in their Table 5, lay far from the cluster proper motion and photometric loci.

Concerning our list of candidate members, after cross matching the lists, we recover 1139 of the candidate members from Table 2 of Stauffer et al. (2007). Seven more than Bouy et al. (2015). Also, we recover the same 34 candidates that Bouy et al. (2015) recovered from the new candidates of Stauffer et al. (2007) (from their Table5).

I have already mentioned that the Table 2 of Stauffer et al. (2007) is an exhaustive compilation of Pleiades members. This list contains objects that authors from the literature classified as Pleiades members even when the membership probability was as low as 0.1 (Stauffer et al. 2007). Therefore, I will not analyse in detail the ~ 300 rejected objects. It suffices to mention that we recover an slightly larger (seven) number of candidates than Bouy et al. (2015).

On the other hand, the list of new candidate members of Stauffer et al. (2007) deserves further attention and a more detailed comparison. From the 21 rejected objects, 14 of them lay below the cluster photometric sequence and far from the proper motion locus. The remaining seven have membership probabilities near our probability threshold. XXXXX
CHECK THIS XXXX

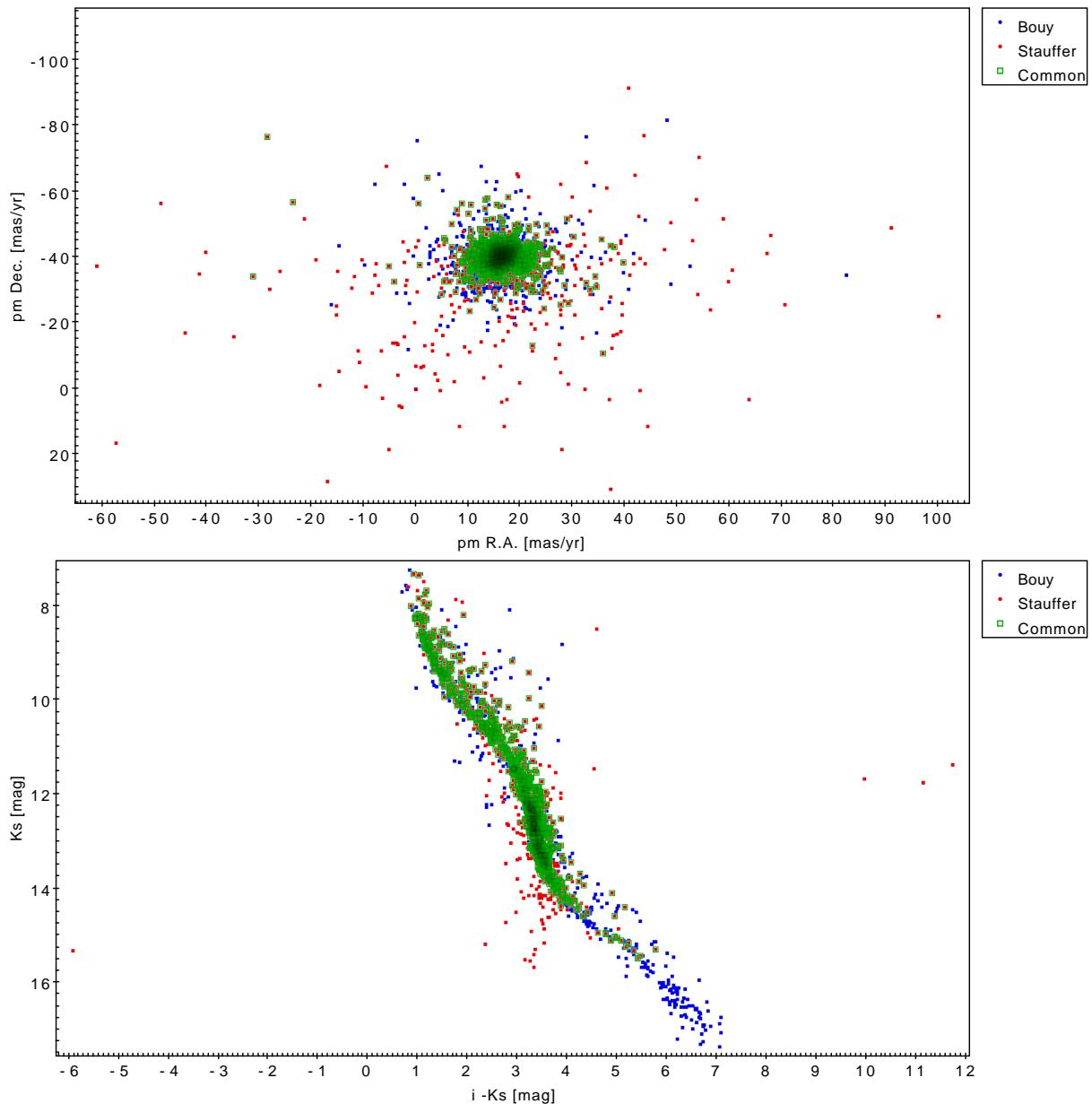


Figure 4.5: Comparison between proper motions (top) and K vs $i - K$ CMD of the candidate members compiled by Stauffer et al. (2007) (red) and those of Bouy et al. (2015) (blue), the common candidates are shown within green squares.

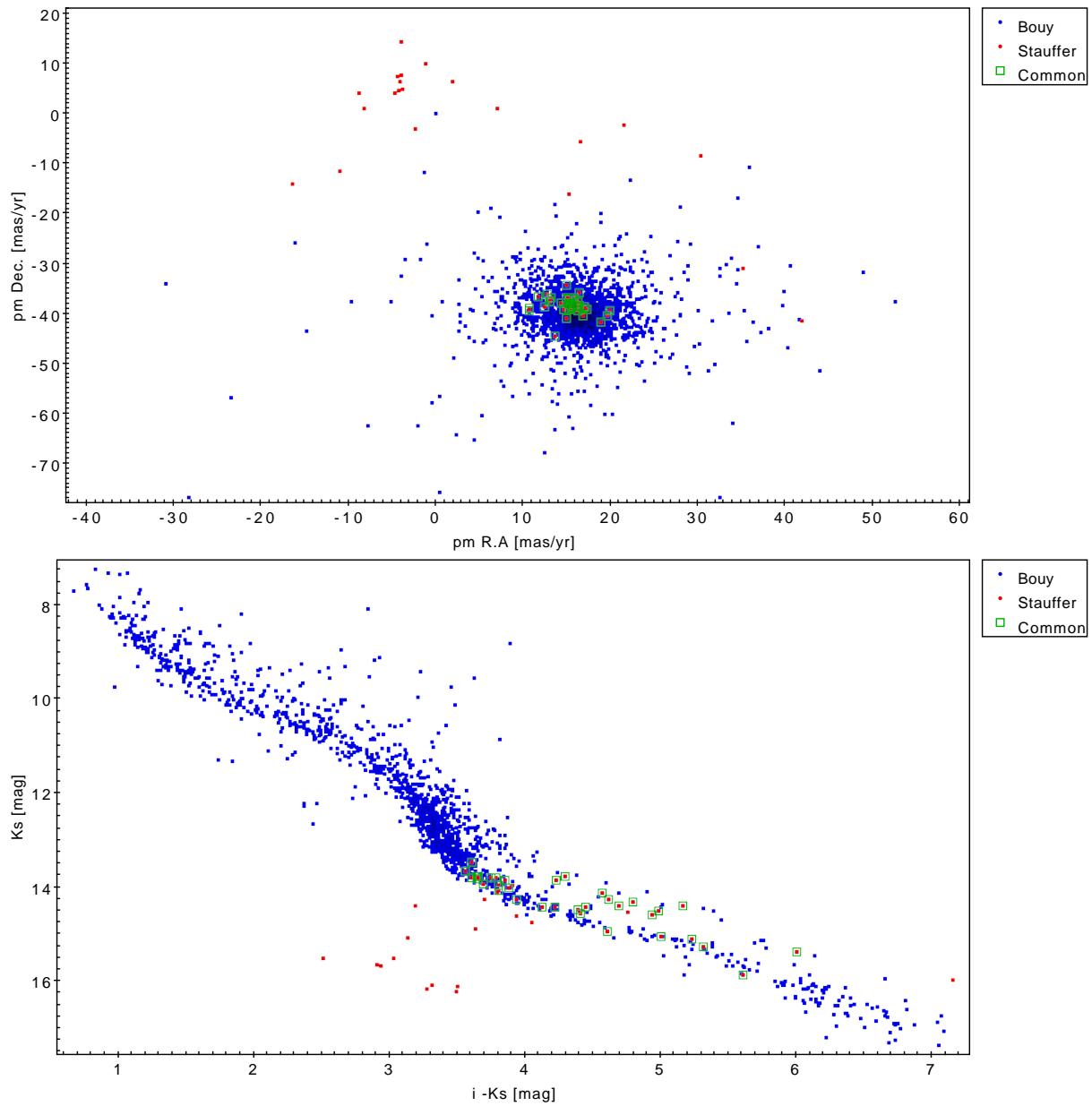


Figure 4.6: Comparison between proper motions (top) and K vs $i - K$ CMD of the new candidate members of Stauffer et al. (2007) (red) and those of Bouy et al. (2015)(blue), the common candidates are shown within green squares.

4.2.2 Candidate members from Bouy et al. (2015)

The methodology of this work, although essentially different from that of Bouy et al. (2015), the fact that both share the data set and use the same observables, allows a direct comparison between them. When using their objective probability thresholds, as shown by Fig. 4.7, both methodologies agree on the outstanding 99.6% of the classified objects. Concerning the candidate members, they also agree on $\approx 90\%$ of them. Nevertheless, the discrepancies are worthy of discussion.

The rejected candidates of Bouy et al. (2015) (lower right box of Fig. 4.7) amount to 12% of their total number of candidate members. This value is 4.7% higher than the contamination rate reported by Sarro et al. (2014): $7.3 \pm 1.4\%$. I will not reject a priori these objects as candidate members because, as mentioned in Sect. 4.1, our methodology recovers 4% less members when compared to that of Sarro et al. (2014)(see Fig. ??). The majority, 85%, of these objects have missing values, which indicates that these objects require future follow up. Even the 37 completely observed and rejected candidates can not be discarded as true members due to the fact that our methodology losses 6% of the true cluster members (see Fig. ??).

CHANGE THIS PARAGRAPH. THE SAMPLES ARE NOT THE SAME THUS 10% IS MEANINGLESS. On the other hand, our new candidates (upper left box of Fig. 4.7) amount to 10% of Bouy et al. (2015) candidates. This figure is higher than the $4.3 \pm 0.2\%$ CR reported on the previous Section. Thus, it indicates that up to $\sim 6\%$ of these objects can be true cluster members. From these objects, one half have completely observed values, thus 3% of our new candidates are completely observed objects. This last figure agrees perfectly well with the 3% of non recovered members reported by Sarro et al. (2014).

In Figs. 4.8 and 4.9 I show the proper motions and K_s vs $i - K_s$ CMD projection spaces of our new candidates and the rejected ones of Bouy et al. (2015). In the following I elaborate on their properties.

UNITS WITHIN MATHEMATICAL EXPRESSION.

MEDIAN SYMBOL? SUBINICES? CHANGE ALSO IN THE PAPER.

The new candidate members have proper motions uncertainties whose median, $\overline{\mu_{\alpha,\delta}} = \{1.33, 1.33\} \text{ mas/yr}$, is two times larger than those of the candidate members in common with Bouy et al. (2015), median $\overline{\mu_{\alpha,\delta}} = \{0.65, 0.65\} \text{ mas/yr}$. Also, as shown by Fig. 4.8, the majority of the new candidate members, 148, have probabilities lower than 0.95, are located in a halo around the locus of the cluster proper motions, and on top of the cluster sequence in the K_s vs $i - K_s$ CMD. On the contrary, the new candidates with probabilities higher than 0.95, which are 39, lay in the centre of the cluster proper motions and fall above the cluster sequence in the K_s vs $i - K_s$ CMD. Thus, I hypothesise that: i) Objects whose photometry is compatible with the cluster sequence but are in the proper motions halo, have higher membership probabilities in our methodology due to the increased flexibility of the cluster proper motions model: it now has four gaussians instead of the

two of Bouy et al. (2015). And ii) objects near the centre of the cluster proper motions but located above the cluster photometric sequence sequence, are multiple systems (probably triple systems which can amount to 4% of the population Duquennoy & Mayor 1991) with an increased membership probability due to our more flexible photometric model of the cluster and equal-mass binaries sequences.

The rejected candidates of Bouy et al. (2015), as it is shown in Figs. 4.9 and 4.10, have proper motions uncertainties with median $\overline{\mu_{\alpha,\delta}} = \{3.15, 3.19\} \text{ mas/yr}$. This value is more than four times larger than that of the candidates in common with our members. Also, these objects are distributed along the cluster sequence. Among these objects, those with a relatively high membership probability occur mostly at the middle of the cluster sequence (green squares of Fig. 4.10) while those with lower membership probabilities occur at the bright and faint ends (blue and red triangles of Fig. 4.10, respectively). These last regions coincide with those where the missing values happen the most. We stress the fact that Sarro et al. (2014) and later Bouy et al. (2015) construct their models using only completely observed objects (i.e. those without missing values). For their field models both authors use a sample of $\approx 20,000$ objects. Proceeding in that way, as explained in Sect. 3.3.1, underestimates the field density, particularly in the regions where the missing values are more frequent. Underestimating the field likelihood increases in the cluster field likelihood ratio, therefore it increases also the cluster membership probabilities. Furthermore, the proper motions uncertainties of objects at the bright middle and faint ends, have medians of $\overline{\mu_{\alpha,\delta}} = \{4.0, 4.2\} \text{ mas/yr}$, $\overline{\mu_{\alpha,\delta}} = \{2.4, 2.4\} \text{ mas/yr}$ and $\overline{\mu_{\alpha,\delta}} = \{3.4, 3.4\} \text{ mas/yr}$, respectively. These figures are approximately 6, 4 and 5 times larger, respectively, than those of the candidates in common. These large uncertainties produce a proportional spread of the likelihood distribution thus further reducing the membership probability.

I consider that both the large proper motion uncertainties and field likelihoods are responsible for the diminished membership probabilities of Bouy et al. (2015) rejected candidates. However, these rejected candidate members cannot be discarded as potential members. Indeed, at the probability threshold of maximum accuracy, $p_t = 0.84$, the TPR is just $90.0 \pm 0.05\%$. It means that there are still 10% of true members within the rejected candidates. To rule out the possibility that these objects are indeed members we need lower proper motion uncertainties and fewer missing values. Future steps will be taken to try to solve this issue.

Summarising, the discrepancies between our individual membership probabilities and those reported by Bouy et al. (2015) arise from subtle but important differences. The first of them is the more formal treatment of missing values in our methodology and its inclusion in the field model. Taking into account the missing values has two main consequences. The first of them is that, the new photometric model of the field diminished membership probabilities, particularly in the regions where missing values happen the most. Second, the use of missing values in the construction of the cluster model allow us

to include the information of good candidate members that were otherwise discarded a priori. The second difference is the higher flexibility of our cluster model, it allows us to increase the membership probability of the previously discarded candidates. Furthermore, as shown by the red squares in the upper left corner of Fig. 4.7, the higher flexibility of our cluster model allow us to include as new candidate members previously rejected objects with complete (non-missing) values.

Figure 4.7: Recovered membership probabilities compared to those of [Bouy et al. \(2015\)](#). Lines show the 0.75 and $p_t = 0.84$ probability thresholds used in both works. The numbers indicate the new candidate members (top left), rejected candidate members (bottom right), and common candidate members (top right).

Figure 4.8: Proper motion (left) and K_s vs. $i - K_s$ CMD (right) showing the new candidate members found in this work. XXXXX.

Figure 4.9: Proper motion (left) and K_s vs. $i - K_s$ CMD (right) showing the rejected candidate members of [Bouy et al. \(2015\)](#). Captions as in Fig. 4.8.

4.3 The statistical distributions of the Pleiades cluster.

Once the objective probability threshold has been established and the results of the classification analysed and compared with the literature. Now I present the results of the statistical distributions that describe the cluster population. These distributions result directly or indirectly from the posterior distribution of the parameters in our model. Here I summarise these parametric posterior distributions, together with some of their correlations. Also, I present an analysis on the way these posterior distribution help us to update our prior knowledge.

In Table ?? I summarise the posterior distribution of the parameters in our model. I use as statistic and uncertainty the mode and the 16th and 84th percentiles, respectively. The parameter names correspond to those given in Sect. 3.4.

4.4 Updating the prior knowledge

As mentioned by [Gelman \(2006\)](#), the posterior distribution must be inspected to update our previous knowledge. To inspect these posterior distribution, I use the bare values

Figure 4.10: Proper motion (left) and K_s vs. $i - K_s$ CMD (right) showing the rejected candidate members of [Bouy et al. \(2015\)](#). The colours and shapes are a proxy for their K_s magnitude.

and the correlations among them. The values indicate, for example, that the number of GMM modelling the proper motions of the single stars is overestimated. The fraction and variance of the last gaussian went both to near zero values. A better model would be that in which no computing power will be lost in inferring negligible parameters.

Another simple inspection tells us that our prior for the cluster field fraction is narrow and its mode lay far from that of the posterior distribution. Although this is not wrong since the prior clearly allow this value, this prior must be made wider. Clearly our prior knowledge underestimated the expected number of cluster members.

Another useful tool to inspect these posterior distributions is the correlation they show with themselves. In Figure 4.11 I show the correlation matrix of all the parameters in our model. This Fig. shows elements that could be considered to improve our model. For example, the large correlation shown by the high order coefficients of the spline series may indicate that we can save some of these parameters.

Figure 4.11: Correlation matrix of parameters in our model. The colour code indicates the value of the correlation coefficient.

In the following Sections, I use the distributions of the parameters in our model to derive the statistical distributions that describe the cluster population. Some of these distributions have as parameters those inferred in the model, however, other, like the mass distribution require more elaborated derivations.

Discuss how to deal with future clusters in which we will not have Bouy2015 for the priors.

IS BIC a good approximation, what can be done in the future?

4.5 Velocity distribution

I must work on it. Here I must transform the bivariate proper motions distributions into a univariate velocity distributions using the distance and assuming spherical symmetry. This distribution will give us some hints to improve the proper motions model. It will be interesting to fit a model, maxwellian for example and see how well it fits. If more models are available we can try to find the best one in terms of Bayesian model selection.

4.6 Spatial distribution

To include

4.7 Luminosity distribution

This Section describes the process by which the apparent J, H , and K_s magnitude distributions are obtained and then transformed into the absolute magnitude distributions. Later, I present these distributions and compare them with those found by Bouy et al. (2015). Also I will compare them with the ones resulted from only the high membership probability objects.

4.7.1 Derivation of the magnitude distributions

To derive the J, H, K_s magnitude distributions I start with the colour index, CI , distribution. This last one, is described by an univariate GMM whose parameters are inferred in the model. Since we also model the EMB, their fraction and photometric sequence are taken into account.

In the following, I exemplify the process of derivation on the K_s band. Similar transformations apply to the rest of the bands. To obtain the distribution of K_s for the cluster objects, I take the colour index CI as a nuisance parameter, later I marginalise it. Thus, REMOVE CI FROM PCI IT IS OBLIVIOUS.

MARGINALISATION OF CI OVER CI OBSERVED.

$$p(K_s|\boldsymbol{\theta}_c) = \int p(K_s, CI|\boldsymbol{\theta}_c) \cdot dCI = \int p(K_s|CI, \boldsymbol{\theta}_c) \cdot p(CI|\boldsymbol{\theta}_c) \cdot dCI.$$

The term $p(K_s|CI, \boldsymbol{\theta}_c)$ corresponds to the GMM modelling the distribution of CI (Eq. 3.28), while $p(K_s|CI, \boldsymbol{\theta}_c)$ is the probability of K_s given the CI , and the cluster parameters $\boldsymbol{\theta}_c$. The EMB are included with an amplitude equal to their fraction, $(1 - \pi_{CB})$. Thus,

$$\begin{aligned} p(K_s|\boldsymbol{\theta}_c) &= \int [\pi_{CB} \cdot p_{Cs}(K_s|CI, \boldsymbol{\theta}_c) + (1 - \pi_{CB}) \cdot p_{Bs}(K_s|CI, \boldsymbol{\theta}_c)] \\ &\quad \cdot p_{CI}(CI|\boldsymbol{\theta}_c) \cdot dCI. \\ &= \pi_{CB} \int p_{Cs}(K_s|CI, \boldsymbol{\theta}_c) \cdot p_{CI}(CI|\boldsymbol{\theta}_c) dCI \\ &\quad + (1 - \pi_{CB}) \int p_{Bs}(K_s|CI, \boldsymbol{\theta}_c) \cdot p_{CI}(CI|\boldsymbol{\theta}_c) \cdot dCI. \end{aligned} \tag{4.1}$$

In this equation, Cs and Bs stand for cluster and EMB sequences, respectively.

The terms inside the integrals correspond to Equations 3.27 and 3.28. However, since here I want to derive only the distribution of K_s , I marginalise the rest of the bands. Also, the integration limits must change to those of the truncated colour distribution ($CI_{min} = 0.8, CI_{max} = 8$). Hence,

$$\begin{aligned} p(K_s | \boldsymbol{\theta}_c) &= \pi_{CB} \int_{CI_{min}}^{CI_{max}} \left[\left[\sum_{i=1}^5 \pi_{CI,i} \cdot \mathcal{N}_t(CI | \mu_{CI,i}, \sigma_{CI,i}) \right] \right. \\ &\quad \cdot \int_{\tilde{Y}, \tilde{J}, \tilde{H}} \mathcal{N}(\{CI, \tilde{Y}, \tilde{J}, \tilde{H}, K_s\} | \boldsymbol{\mathcal{S}}(CI, \boldsymbol{\beta}), \Sigma_{clus}) d\tilde{Y} d\tilde{J} d\tilde{H} \Big] \cdot dCI \\ &\quad + (1 - \pi_{CB}) \int_{CI_{min}}^{CI_{max}} \left[\left[\sum_{i=1}^5 \pi_{CI,i} \cdot \mathcal{N}_t(CI | \mu_{CI,i}, \sigma_{CI,i}) \right] \right. \\ &\quad \cdot \int_{\tilde{Y}, \tilde{J}, \tilde{H}} \mathcal{N}(\{CI, \tilde{Y}, \tilde{J}, \tilde{H}, K_s\} | T_{Bs}(\boldsymbol{\mathcal{S}}(CI, \boldsymbol{\beta})), \Sigma_{clus}) d\tilde{Y} d\tilde{J} d\tilde{H} \Big] \cdot dCI. \end{aligned}$$

The derivations of the J and H magnitude distributions are similar to the procedure described for K_s . It is worth of notice that, the derivation process takes into account the EMB and the systems (binaries or multiples) which could have different mass ratios. Therefore, we call them the system magnitude distributions.

The previous magnitude distributions, together with the parallax and extinction of the cluster, allows us to obtain the system luminosity distributions. Here, I assume that the parallax is normally distributed with mean, 7.44 mas, and standard deviation 0.42 mas (Galli et al. 2017). This parallax distribution is then convolved with the magnitude distributions to obtain the absolute magnitude distributions. Finally, I deredden them employing the canonical value of extinction: $A_v = 0.12$ (Guthrie 1987). This last values were transformed into the J, H, K_s extinctions using the extinction law of Cardelli et al. (1989).

Our methodology prescribes the *true* photometric quantities based on the *true* colour index CI . Therefore, the completeness limits of this CI dictate those of the photometric bands. The upper completeness limits that Bouy et al. (2015) estimate for i and K_s are $i \approx 23$ mag and $K_s \approx 18$ mag (see their appendix A). As these authors mention, due to the heterogeneous origins of the DANCE DR2 survey, the completeness is not homogeneous over its entire area. To overcome this issue, they identified a region, the inner three degrees of the cluster, with homogeneous spatial and depth coverage. Then, they restricted their analysis to this region.

Restricting the sample means that good candidate members are not taken into account. Furthermore, if any dynamical process has been set on the cluster such that the mass distribution of its members is not uniform in the space, then restricting the sample could introduce a bias. One of such dynamical process is mass segregation, which, as suggested by Adams et al. (2001) may have happen in the Pleiades.

Instead of restricting the sample, I assume that the UKIDSS survey, which is the most profound of the DANCe catalogue, provides the homogeneous spatial coverage at faint magnitudes, thus providing the lower limit to the completeness. Then, for the bright end of the survey, I quote conservative completeness limits. Figure 4.12 shows the K_s and i density for all sources in the Pleiades DANCe DR2. The upper completeness limits correspond to the point with maximum density, $i = 21.4$ mag, $K_s = 18.1$ mag. The density at bright magnitudes shows a sharp decline, probably due to saturation. Therefore, I choose $i = 13.2$ mag and $K_s = 11.0$ mag as the lower completeness limits.

The CI completeness interval is then defined as that of all the points, along the cluster sequence in the K_s vs. $i - K_s$ CMD, for which i and K_s are bounded by their upper and lower completeness limits, respectively. This results in a completeness interval for $i - K_s$ of $2.7 < i - K_s < 5.6$ mag. With this completeness interval and the cluster sequence (the splines), we derive the completeness intervals for the J, H, K_s . Finally, these intervals were transformed to absolute magnitudes and dereddened.

Figure 4.12: Density of all DANCe DR2 sources in K_s vs i magnitudes. Lines show our completeness limits, $13.2 < i < 21.4$ mag and $11 < K_s < 18.1$ mag. The grey area is considered incomplete.

In Fig. 4.13, I plot the J, H, K_s luminosity distributions together with their completeness limits, hereafter I call these distributions the continuous BHM. For the sake of comparison I also show the following luminosity distributions. The luminosity distributions of objects whose mode of membership probability is greater than our probability threshold $p_t = 0.84$, I call these distributions the discrete BHM. Also, I plot the luminosity distribution resulting from the candidate members of Bouy et al. (2015), I call it discrete Bouy. Since the discrete luminosity distributions, both Bouy and BHM, relay on the individual object magnitudes, and many of these objects have missing values, I impute them using the nearest euclidean neighbour.

The difference between the continuous BHM and discrete BHM distributions comes essentially from the objects used to obtain them. The continuous one uses all objects proportionally to their cluster membership probability while the discrete BHM uses only the high probability candidate members ($p > p_t$). Since the discrete BHM is not a random sample of the continuous BHM, therefore their distributions does not need to be exactly alike. In addition, the missing values in the continuous BHM case were marginalised, while in the discrete BHM were imputed.

On the other hand, the differences between the discrete distributions, BHM and that of Bouy et al. (2015), arise mainly at the bright and faint ends ($K_s \approx 4$ mag and $K_s \approx 11$ mag). We argue that the origin of these differences lay in our new candidate members and in the rejected ones of Bouy et al. (2015), as it is discussed in Sect. 4.2.2.

Figure 4.13: Luminosity functions from J, H, K_s (orange). Also shown the regions of incompleteness and the luminosity functions computed from: the candidate members of [Bouy et al. \(2015\)](#) (dot-dashed blue line), and our candidate members, ($p_{84\%} > p_t$, dashed black line).

4.8 Mass distribution

This Section starts with a brief description of the mass-luminosity relation, which is used to derive the present day system mass function (PDSMF) of the Pleiades together with the luminosity distribution of the previous section. Then I compare this PDSMF to the Initial Mass Functions (IMFs) of [Chabrier \(2005\)](#) and [Thies & Kroupa \(2007\)](#). Later, I conclude this section with an analysis of some simple models that can be fitted to our the PDSMF, I give the best model according to the Bayesian evidence.

4.8.1 The mass-luminosity relation

The mass-luminosity relation is the key to transform luminosities into masses. This relation relies entirely on the current models of stellar atmosphere and evolution. Among the different flavours of models, we choose the BT-Settl model of [Allard et al. \(2012\)](#). We based this decision on the fact that currently, this is the only model which covers our luminosity range at the age of the Pleiades. Since the DANCe survey uses as reference the J, H and K_s bands of the 2MASS survey, I choose the BT-Settle grid: CIFIST2011bc for the 2MASS AB photometric system. This grid, as it name indicates, returns values of the luminosity on a (non-uniform) grid on the mass. However, the transformation of luminosities into masses is proportional and thus very sensitive to the Jacobian of the transformation. To avoid the discontinuities in the derivatives produced by the grid, we decided to fit the grid using spline series and then obtain the derivative from these series. It is important to notice that we implicitly assume that the transformation from luminosities to masses does not have any associated uncertainty, thus it is uniquely determined. We can not do more since the models do not provide any uncertainty.

4.8.2 Present day mass function

The PDSMF is independently obtained in the J, H, K_s bands by transforming the luminosity functions into system mass functions using the mass-luminosity relations described in the previous section. Since the luminosity functions of Sect. 4.7 correspond to the luminosity of systems (single and binary stars altogether), therefore the derived mass function corresponds to the system mass function. We assume an age of 120 Myr for the Pleiades together with solar metallicity, these are the canonical values reported by and , respectively.

Figure 4.14 shows the logarithmic PDSMF (ξ_L) for the J, H, K_s bands normalised on the completeness limits obtained in Sect. 4.7. This figure also shows, the PDSMF proposed by Bouy et al. (2015) and, the IMFs of Thies & Kroupa (2007) and Chabrier (2005). For this last IMF, I show the standard uncertainty as the value reported by Chabrier (2003). As shown in this Fig., our PDSMFs compares well, in the completeness interval, with the one proposed by Bouy et al. (2015). However, there are discrepancies, particularly above $0.3M_\odot$ ($-0.5 < \log M/M_\odot$). These may have its origin on the higher membership probability of our new candidate members. These new candidate members are preferentially M stars, whose masses are the range $0.075 - 0.6M_\odot$ ($-1.12 < \log M/M_\odot < -0.22$). Furthermore, these differences root also on the objects that Bouy et al. (2015) did not include in their analysis: those lying outside the inner three degree region.

To obtain a model for our PDSMF I proceed as follows. First, I select three models: a log-normal distribution and two power-law distributions of the form $m^{-\alpha}$, with two and three power-law segments. Second, I took the mode distribution of our 100 sample distributions of the PDSMF in the K_s band and completeness interval and draw a 10^4 synthetic sample. Afterwards, using *PyMultiNest* (Buchner et al. 2014), and the synthetic sample I fitted the three models. Finally, in Table 4.1, I give the posterior distributions of each model parameters together with its evidence (see Sect. 3.1.2).

Judging by the evidences, the best model corresponds to the two segment power-law distribution. This model is over plot as solid black line in Fig. 4.14. This model agrees well with that found by Bouy et al. (2015), except for the flat part in the low-mass range and the less steep slope in the high mass range. Nevertheless, it is in clear discrepancy with the IMFs of Chabrier (2005), ($m_c = 0.25_{-0.016}^{+0.021}$ and $\sigma = 0.55_{-0.01}^{+0.05}$, the uncertainties are those reported by Chabrier 2003, for single objects) and of Thies & Kroupa (2007). The discrepancy between the IMFs and the PDSMFs (Bouy et al. 2015, and ours) may have its origin on the not yet established uncertainties in the mass-luminosity relationship, on dynamical effects associated with age, or on both of them. In the next section I compare the PDSMF of the Pleiades with that of other younger and older clusters in order to analyse if there is substantial evidence to claim for dynamical effects associated with age.

Table 4.1: Parameters and evidence of models fitted to the PDSMF

Model	Parameters	Log Evidence
LogNormal	$m_c = 0.36 \pm 0.03$ $\sigma = 0.46 \pm 0.02$	18.1 ± 0.1
Two Segments	$\alpha_0 = -0.11 \pm 0.06 \quad m \in [0.04, 0.22 \pm 0.01]$ $\alpha_1 = 1.13 \pm 0.1 \quad m \in [0.22 \pm 0.01, 0.56]$	2222.7 ± 0.4
Three Segments	$\alpha_0 = -0.05 \pm 0.6 \quad m \in [0.04, 0.08 \pm 0.03]$ $\alpha_1 = -0.1 \pm 0.1 \quad m \in [0.08 \pm 0.03, 0.22 \pm 0.01]$ $\alpha_2 = 1.13 \pm 0.1 \quad m \in [0.22 \pm 0.01, 0.56]$	2221.2 ± 0.3

Before concluding this section, I use our PDSMF to give a lower limit to the mass of the cluster. The cluster members mean mass, in our entire mass range, is $0.26 \pm 0.006 M_{\odot}$. Thus, the product of this mean mass with the expected number of cluster members, 3116 ± 110 , is an estimate of mass of the cluster: $807^{+38}_{-29} M_{\odot}$. The expected number of cluster members is the integral, over the whole range of membership probabilities, of number of objects at each membership probability. However, since we still lack the low and high mass ranges of the PDSMF, this value is only a lower limit to the mass of the cluster. As mentioned in Sect. 4.8.1, the uncertainties in the mass-luminosity relations are yet to be established. Thus the quoted uncertainties of our mass results are underestimated.

Figure 4.14: Normalised PDSMF in J, H, K_s bands. Also shown the IMFs of Chabrier (2005); Thies & Kroupa (2007) and fits to the PDSMF found by us and Bouy et al. (2015).

4.9 The mass distribution on time

In order to test if dynamical effects could be the origin of the discrepancies between the PDSMFs and the IMFs, in this section I analyse the evolution in time of the mass distribution. To do this, I compare the differences between the Pleiades PDSMF (≈ 120 Myr) and those of the Trapezium and Hyades clusters, which are ≈ 1 Myr and ≈ 600 Myr old respectively. These can be thought as snapshots of the Pleiades pasts and future mass functions.

In Figure 4.15 I compare the PDSMF from the Pleiades derived here, to those of the Trapezium and Hyades. These PDSMFs correspond to those of Fig. 11 of Bouy et al. (2015) (private communication). As mentioned by Bouy et al. (2015), the abundance of low-mass stars and brown dwarfs in the range $0.03 - 0.1 M_{\odot} (\log M/M_{\odot} \approx \{-1, -1.4\})$ seems to diminish with time. Furthermore, since the PDSMF are normalised, a diminish in the lower mass range produces a relative increase of low-mass stars in the range $-0.4 < \log M/M_{\odot} < -0.2$. The alternative effect of an increase in the high mass range which then may produce a relative diminish in the low mass range, although is statistically possible, is unlikely in the astrophysical sense. Open clusters tend lose stars, not the other way around. Furthermore, this effect is consistent with the classical scenario in which low-mass stars and brown dwarfs are ejected as the cluster relaxes. I test the validity of this scenario, at least the statistical significance of the observed differences among the PDSMF of this three clusters. To perform this test, the null hypothesis is that the Trapezium and the Hyades have the same PDSMF as the Pleiades. Since I only have the inferred model of the Pleiades cluster, I am compelled to perform a frequentist test. Thus, to do the statistical comparison of these three PDSMF I use the Kolmogorov-Smirnov and the Anderson-Darling tests.

Figure 4.15: Left: PDSMFs of the Pleiades (derived here for J, H, K_s bands), Trapezium, and Hyades, from [Bouy et al. \(2015\)](#). They are normalised in the interval of completeness.

Figure 4.16: Cumulative distribution functions (CDF) of the PDSMFs from left panel and that of [Chabrier \(2005\)](#) and [Thies & Kroupa \(2007\)](#) system initial mass function (normalised also in the interval of completeness). The Pleiades CDF shown is just from K_s band. The grey area depicts the area in which the null hypothesis of same PDSMF as that of the Pleiades can not be rejected (at $\alpha = 0.01$).

Figure 4.16 shows the cumulative distribution functions (CDFs) of the Trapezium, Pleiades (only in K_s band) and Hyades PDSMFs. Also and for comparison, we show the CDFs of [Chabrier \(2005\)](#) and [Thies & Kroupa \(2007\)](#) IMFs. The grey area around the Pleiades CDF shows the hypothesis test in which I compare each CDF with that of the Pleiades. The null hypothesis, as mentioned before is that each compared CDF is exactly that of the Pleiades. I use the Kolmogorov-Smirnov statistic and the alpha value $\alpha = 0.01$, to compute the maximum vertical distance d_α from the Pleiades CDF, the grey region was created with this maximum distance. The null hypothesis is rejected only if the tested CDF lies entirely outside the grey region around the Pleiades CDF. As can be seen, neither the IMFs nor the PDSMF of the Trapezium and Hyades lay entirely within the grey area, thus we can reject the null hypothesis that they share the same PDSMF of the Pleiades. Furthermore, since the Kolmogorov-Smirnov test uses only the maximum distance between CDFs, we also applied the more robust Anderson-Darling test. It also rejects the null hypotheses (at $p < 0.004$) that the Trapezium and Hyades PDSMFs and the [Chabrier \(2005\)](#) and [Thies & Kroupa \(2007\)](#) IMFs have the same CDF of the Pleiades.

The previous tests show that there is enough evidence to claim for differences among the PDSMFs of these three clusters and from IMFs and Pleiades PDSMF. Thus they suggest that the observed differences may have an origin on dynamical effects associated with age and relaxation. Nevertheless, to claim for reliable evidence supporting these differences the census of the Trapezium and Hyades must be done using the same methods. Also, the uncertainties must be properly establish both for the other PDSMFs and for the mass-luminosity relation from which all these PDSMF are derived.

Chapter 5

Conclusions and Future Work

Bibliography

- Adams, J. D., Stauffer, J. R., Monet, D. G., Skrutskie, M. F., & Beichman, C. A. 2001, AJ, 121, 2053
- Adams, W. S. 1904, ApJ, 19
- Akeret, J., Seehars, S., Amara, A., Refregier, A., & Csillaghy, A. 2013, Astronomy and Computing, 2, 27
- Allard, F. 2014, in IAU Symposium, Vol. 299, Exploring the Formation and Evolution of Planetary Systems, ed. M. Booth, B. C. Matthews, & J. R. Graham, 271–272
- Allard, F., Homeier, D., & Freytag, B. 2013, Mem. Soc. Astron. Italiana, 84, 1053
- Allard, F., Homeier, D., Freytag, B., & Sharp, C. M. 2012, in EAS Publications Series, Vol. 57, EAS Publications Series, ed. C. Reylé, C. Charbonnel, & M. Schultheis, 3–43
- Alsing, J., Heavens, A., & Jaffe, A. H. 2017, MNRAS, 466, 3272
- Anderes, E., Wandelt, B. D., & Lavaux, G. 2015, ApJ, 808, 152
- Andrieu, C., de Freitas, N., Doucet, A., & Jordan, M. I. 2003, Machine Learning, 50, 5
- Astraatmadja, T. L. & Bailer-Jones, C. A. L. 2016a, ApJ, 832, 137
- Astraatmadja, T. L. & Bailer-Jones, C. A. L. 2016b, ApJ, 833, 119
- Bailer-Jones, C. A. L. 2015, PASP, 127, 994
- Baraffe, I., Chabrier, G., Allard, F., & Hauschildt, P. H. 1998, A&A, 337, 403
- Barnes, III, T. G., Moffett, T. J., Jefferys, W. H., & Forestell, A. D. 2004, in Astronomical Society of the Pacific Conference Series, Vol. 310, IAU Colloq. 193: Variable Stars in the Local Group, ed. D. W. Kurtz & K. R. Pollard, 95
- Bate, M. R., Bonnell, I. A., & Bromm, V. 2003, MNRAS, 339, 577
- Bishop, C. 2006, Pattern Recognition and Machine Learning, Information Science and Statistics (Springer)

- Blaauw, A. 1964, in IAU Symposium, Vol. 20, The Galaxy and the Magellanic Clouds, ed. F. J. Kerr, 50
- Blackwell, T. & Bentley, P. 2002, in Proceedings of the 4th Annual Conference on Genetic and Evolutionary Computations, ed. W. Langdon, E. Cantu-Paz, K. Mathias, R. Roy, & D. Davis (San Francisco, CA, USA: Morgan kaufmann Publishers Inc.), 19–26
- Bouy, H., Bertin, E., Moraux, E., et al. 2013, A&A, 554, A101
- Bouy, H., Bertin, E., Sarro, L. M., et al. 2015, A&A, 577, A148
- Brooks, S., Gelman, A., Jones, G., & Meng, X. 2011, Handbook of Markov Chain Monte Carlo, Chapman & Hall/CRC Handbooks of Modern Statistical Methods (CRC Press)
- Buchner, J., Georgakakis, A., Nandra, K., et al. 2014, A&A, 564, A125
- Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245
- Carpenter, B., Gelman, A., Hoffman, M., et al. 2017, Journal of Statistical Software, Articles, 76, 1
- Carpenter, J. M. 2000, AJ, 120, 3139
- Chabrier, G. 2003, PASP, 115, 763
- Chabrier, G. 2005, in Astrophysics and Space Science Library, Vol. 327, The Initial Mass Function 50 Years Later, ed. E. Corbelli, F. Palla, & H. Zinnecker, 41
- Chung, Y., Gelman, A., Rabe-Hesketh, S., Liu, J., & Dorie, V. 2015, Journal of Educational and Behavioral Statistics, 40, 136
- Clerc, M. & Kennedy, J. 2002, Trans. Evol. Comp, 6, 58
- Converse, J. M. & Stahler, S. W. 2008, ApJ, 678, 431
- Converse, J. M. & Stahler, S. W. 2010, MNRAS, 405, 666
- De Boor, C. 1978, A practical guide to splines / Carl de Boor (Springer-Verlag New York), xxiv, 392 p. :
- Demory, B.-O. 2014, ApJ, 789, L20
- Dempster, A. P., Laird, N. M., & Rubin, D. B. 1977, JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B, 39, 1
- Diard, J. & Bessiere, P. 2008, Probabilistic reasoning and decision making in sensory-motor systems, 153

- Dirac, P. A. M. 1942, Proceedings of the Royal Society of London Series A, 180, 1
- Duane, S., Kennedy, A. D., Pendleton, B. J., & Roweth, D. 1987, Physics Letters B, 195, 216
- Duquennoy, A. & Mayor, M. 1991, A&A, 248, 485
- Faherty, J. K., Burgasser, A. J., Walter, F. M., et al. 2012, ApJ, 752, 56
- Fan, Y. & Sisson, S. 2011, Handbook of Markov Chain Monte Carlo, 67
- Feeney, S. M., Johnson, M. C., McEwen, J. D., Mortlock, D. J., & Peiris, H. V. 2013, Phys. Rev. D, 88, 043012
- Fei-Fei, L. & Perona, P. 2005, in Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, Vol. 2, Ieee, 524–531
- Feroz, F., Hobson, M. P., & Bridges, M. 2009, Monthly Notices of the Royal Astronomical Society, 398, 1601
- Flegal, J. M., Hughes, J., & Vats, D. 2016, mcmcse: Monte Carlo Standard Errors for MCMC, Riverside, CA and Minneapolis, MN, r package version 1.2-1
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, Publications of the Astronomical Society of the Pacific, 125, 306
- Franklin, J. 2002, The Science of Conjecture: Evidence and Probability Before Pascal, A Johns Hopkins paperback (Johns Hopkins University Press)
- Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1
- Gaia Collaboration, van Leeuwen, F., Vallenari, A., et al. 2017, A&A, 601, A19
- Galli, P. A. B., Moraux, E., Bouy, H., et al. 2017, A&A, 598, A48
- Gatewood, G., de Jonge, J. K., & Han, I. 2000, ApJ, 533, 938
- Gelman, A. 2006, Bayesian Analysis, 1, 515
- Gelman, A. 2012, CHANCE, 25, 52
- Gelman, A., Carlin, J., Stern, H., et al. 2013, Bayesian Data Analysis, third edition edn. (CRC)
- Gelman, A. & Hill, J. 2007, Data Analysis Using Regression and Multilevel/Hierarchical Models, Analytical Methods for Social Research (Cambridge University Press)
- Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. 2008, Ann. Appl. Stat., 2, 1360

- Geman, S. & Geman, D. 1984, IEEE Trans. Pattern Anal. Mach. Intell., 6, 721
- Gong, L. & Flegal, J. M. 2016, Journal of Computational and Graphical Statistics, 25, 684
- Good, I. J. 1980, Trabajos de Estadistica Y de Investigacion Operativa, 31, 489
- Goodman, J. & Weare, J. 2010, Communications in Applied Mathematics and Computational Science, 5, 65
- Guthrie, B. N. G. 1987, QJRAS, 28, 289
- Hambly, N. C. & Jameson, R. F. 1991, MNRAS, 249, 137
- Hastings, W. K. 1970, Biometrika, 57, 97
- Hertzsprung, E. 1947, Annalen van de Sterrewacht te Leiden, 19, A1
- Hodgkin, S. T. & Jameson, R. F. 2000, in Astronomical Society of the Pacific Conference Series, Vol. 198, Stellar Clusters and Associations: Convection, Rotation, and Dynamos, ed. R. Pallavicini, G. Micela, & S. Sciortino, 59
- Hogg, D. W., Bovy, J., & Lang, D. 2010a, ArXiv e-prints
- Hogg, D. W., Myers, A. D., & Bovy, J. 2010b, ApJ, 725, 2166
- Huang, A. & Wand, M. P. 2013, Bayesian Analysis, 8, 439
- Jameson, R. F., Dobbie, P. D., Hodgkin, S. T., & Pinfield, D. J. 2002, MNRAS, 335, 853
- Jaynes, E. 2003, Probability Theory: The Logic of Science (Cambridge University Press)
- JCGM. 2008, JCGM 100: Evaluation of Measurement Data - Guide to the Expression of Uncertainty in Measurement, Tech. rep., Joint Committee for Guides in Metrology
- Jefferys, T. R., Jefferys, W. H., Barnes, III, T. G., & Dambis, A. 2007, in Astronomical Society of the Pacific Conference Series, Vol. 371, Statistical Challenges in Modern Astronomy IV, ed. G. J. Babu & E. D. Feigelson, 433
- Johnson, H. L. & Mitchell, R. I. 1958, ApJ, 128, 31
- Jones, B. F. 1970, AJ, 75, 563
- Kass, R. E. & Raftery, A. E. 1995, Journal of the American Statistical Association, 90, 773
- Kennedy, J. & Eberhart, R. 1995, in Neural Networks, 1995. Proceedings., IEEE International Conference on, Vol. 4, 1942–1948 vol.4
- King, I. 1962, AJ, 67, 471

- Koller, D. & Friedman, N. 2009, Probabilistic Graphical Models: Principles and Techniques, Adaptive computation and machine learning (MIT Press)
- Lada, C. J. & Lada, E. A. 2003, *ARA&A*, 41, 57
- Lee, S.-W. & Sung, H. 1995, *Journal of Korean Astronomical Society*, 28, 45
- Limber, D. N. 1961, *ApJ*, 134, 537
- Limber, I. D. N. 1962, *ApJ*, 135, 16
- Lin, T., Lee, J., & Ho, H. 2006, WOS:000236696400015
- Lindstrom, M. 1999, *Journal of Computational and Graphical Statistics*, 8, 333
- Lodieu, N., Deacon, N. R., & Hambly, N. C. 2012, *MNRAS*, 422, 1495
- Lodieu, N., Dobbie, P. D., Deacon, N. R., et al. 2007, *MNRAS*, 380, 712
- Loktin, A. V. 2006, *Astronomy Reports*, 50, 714
- Lutz, T. E. & Kelker, D. H. 1973, *PASP*, 85, 573
- March, M. C., Karpenka, N. V., Feroz, F., & Hobson, M. P. 2014, *MNRAS*, 437, 3298
- McMichael, D. W. 1996, in Proc. Fourth International Symposium on Signal Processing and its Applications (ISSPA, 377–378
- Melis, C., Reid, M. J., Mioduszewski, A. J., Stauffer, J. R., & Bower, G. C. 2014, *Science*, 345, 1029
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. 1953, *The Journal of Chemical Physics*, 21, 1087
- Moraux, E., Bouvier, J., Stauffer, J. R., & Cuillandre, J.-C. 2003, *A&A*, 400, 891
- Moraux, E., Kroupa, P., & Bouvier, J. 2004, *A&A*, 426, 75
- Morris, C. N. 1983, *Journal of the American Statistical Association*, 78, 47
- Neal, R. M. 1996, Bayesian Learning for Neural Networks (Secaucus, NJ, USA: Springer-Verlag New York, Inc.)
- Olivares, J., Sarro, L., Moraux, E., Berihuete, A., & Bouy, H. 2017, å, submitted
- Patel, P. K., Sharma, V., & Gupta, K. 2013, *International Journal of Computer Applications*, 73
- Perryman, M. A. C., Lindegren, L., Kovalevsky, J., et al. 1997, *A&A*, 323, L49

- Pinfield, D. J., Jameson, R. F., & Hodgkin, S. T. 1998, MNRAS, 299, 955
- Porras, A., Christopher, M., Allen, L., et al. 2003, AJ, 126, 1916
- Press, W. H. 1997, in Unsolved Problems in Astrophysics, ed. J. N. Bahcall & J. P. Ostriker, 49–60
- Pritchard, C. 1884, MNRAS, 44, 355
- Raboud, D. & Mermilliod, J.-C. 1998, A&A, 329, 101
- Roberts, G. O. & Rosenthal, J. S. 2004, Probability Surveys, 1, 20
- Sale, S. E. 2012, MNRAS, 427, 2119
- Sarro, L. M., Bouy, H., Berihuete, A., et al. 2014, Astronomy & Astrophysics, 14
- Schwarz, G. 1978, Ann. Statist., 6, 461
- Shariff, H., Dhawan, S., Jiao, X., et al. 2016, MNRAS, 463, 4311
- Shkedy, Z., Decin, L., Molenberghs, G., & Aerts, C. 2007, MNRAS, 377, 120
- Skilling, J. 2004, in American Institute of Physics Conference Series, Vol. 735, American Institute of Physics Conference Series, ed. R. Fischer, R. Preuss, & U. V. Toussaint, 395–405
- Skilling, J. 2006, Bayesian Anal., 1, 833
- Soderblom, D. R., Nelan, E., Benedict, G. F., et al. 2005, AJ, 129, 1616
- Spiriti, S., Eubank, R., Smith, P. W., & Young, D. 2013, Journal of Statistical Computation and Simulation, 83, 1020
- Stauffer, J. R., Hartmann, L. W., Fazio, G. G., et al. 2007, ApJS, 172, 663
- Stauffer, J. R., Schultz, G., & Kirkpatrick, J. D. 1998, ApJ, 499, L199
- Tapiador, D., Berihuete, A., Sarro, L. M., Julbe, F., & Huedo, E. 2017, Astronomy and Computing, 19, 1
- Thies, I. & Kroupa, P. 2007, ApJ, 671, 767
- Titus, J. 1938, AJ, 47, 25
- Trumpler, R. J. 1921, Lick Observatory Bulletin, 10, 110
- van Leeuwen, F. 1980, in IAU Symposium, Vol. 85, Star Clusters, ed. J. E. Hesser, 157–162
- van Leeuwen, F. 1999, A&A, 341, L71

- van Leeuwen, F. 2009, A&A, 497, 209
- Wolfgang, A. & Lopez, E. 2015, ApJ, 806, 183
- Woolley, R. V. D. R. 1956, MNRAS, 116, 296